

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 03:07:05 ; Search time 95 Seconds  
(without alignments)  
1126.574 Million cell updates/sec

Title: AAH26175  
Perfect score: 1084  
Sequence: 1 MKFKLVNSARQYKDLNMS.....PSPTSTVTITQAPSSNRPIV 209

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 03: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1084	100.0	209	Q8TAR1	O8tar1 homo sapien
2	984	90.8	234	Q8BN75	Q8bn75 mus musculus
3	984	90.8	420	Q8CBR9	Q8cbr9 mus musculus
4	984	90.8	487	1 ATF2 HUMAN	P15336 homo sapien
5	984	90.8	487	1 ATF2 MOUSE	P15336 mus musculus
6	984	90.8	487	2 Q8CGB4	Q8cgb4 mus musculus
7	980	90.4	487	1 ATF2 RAT	Q00969 rattus norv
8	960	88.6	487	1 ATF2 CHICK	Q93602 gallus gall
9	923	85.1	486	2 Q91576	Q91576 xenopus lae
10	719	66.3	440	2 Q68FE3	Q68fe3 mus musculus
11	571	52.7	413	2 Q8R0S1	Q8r0s1 mus musculus
12	555.5	51.2	494	1 ATF7 HUMAN	P17544 homo sapien
13	497	45.8	307	2 Q9UD83	Q9ud83 homo sapien
14	417	38.5	117	2 Q8IVR8	Q8ivr8 homo sapien
15	398.5	36.8	148	2 Q75N02	Q75n02 homo sapien
16	390.5	36.0	508	1 CRB5 HUMAN	Q02930 homo sapien
17	317.5	29.3	135	2 Q8BKX9	Q8bkx9 mus musculus
18	317.5	29.3	144	2 Q8BM42	Q8bm42 m mus muscu
19	208	19.2	144	2 Q96J78	Q96j78 homo sapien
20	167.5	15.5	850	2 Q9W0Z5	Q9w0z5 drosophila
21	133.5	12.3	593	2 Q8QGH4	Q8qgh4 brachydanio
22	122.5	11.3	503	2 Q9NAR8	Q9nar8 brachioisto
23	116.5	10.7	775	1 Z515 HUMAN	Q9neae6 homo sapien
24	115	10.6	183	2 Q9PVD6	Q9pvd6 brachydanio
25	115	10.6	442	2 Q93311	Q93311 brachydanio
26	115	10.6	565	1 SCRL1 SCHPO	O14335 schizosacch
27	114.5	10.6	454	2 Q6CGR7	Q6cgr7 yarrowia li
28	114	10.5	780	2 Q9PWQ1	Q9pwq1 fugu rubrip
29	114	10.5	780	2 Q9YGM3	Q9ygm3 fugu rubrip
30	112	10.3	447	1 Z1C1 HUMAN	O15915 homo sapien
31	110.5	10.2	1367	1 AMYH_YEAST	P08640 saccharomyc

32	110.5	10.2	1367	2	Q6LCS8	Q6lcs8 saccharomyc
33	110.5	10.2	1426	2	Q7QJK9	Q7qjk9 anopheles g
34	110	10.1	447	1	Z1C1 MOUSE	P45684 mus musculus
35	110	10.1	447	2	Q6PAK5	Q6pak5 mus musculus
36	110	10.1	447	2	Q9JKY2	Q9jky2 rattus norv
37	110	10.1	447	2	Q8OY18	Q8oy18 mus musculus
38	110	10.1	453	2	Q8OY18	Q8oy18 mus musculus
39	109.5	10.1	881	2	Q89Z25	Q89z25 mus musculus
40	109.5	10.1	428	1	RSV1 SCHPO	Q9pvd9 schizosacch
41	109.5	10.1	674	2	Q8OW63	Q8ow63 mus musculus
42	109.5	10.1	675	1	MTF1 MOUSE	Q07243 mus musculus
43	109	10.1	675	2	Q9JTW8	Q9jtw8 mus musculus
44	109	10.1	568	2	Q8BMU0	Q8bmu0 mus musculus
45	109	10.1	791	2	Q9GRX2	Q9grx2 drosophila
			791	2	Q9NFS1	Q9nfs1 drosophila

ALIGNMENTS

RESULT 1					
Q8TAR1					
ID	Q8TAR1	PRELIMINARY;	PRT;	209 AA.	
AC	Q8TAR1;				
DT	01-JUN-2002 (TrEMBLrel. 21, Created)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				
DE	ATF2 protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
NCBI_TaxID=9606;					
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Skin;				
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,				
RA	Jones S.J., Marra M.A.;				
RA	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
PL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Skin;				
RA	Strausberg R.;				
RA	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC026175; AAH26175.1; -				
DR	HSSP; P15336; 1BHI.				
DR	GO; GO:0005634; C:nucleus; IEA.				
DR	GO; GO:0003676; F:nucleic acid binding; IEA.				
DR	GO; GO:0008270; F:zinc ion binding; IEA.				
DR	InterPro; IPR007087; Znf_C2H2.				
DR	Pfam; PF00096; zf-C2H2; 1.				
DR	SMART; SM00355; Znf_C2H2; 1.				
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.				
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.				
SQ	SEQUENCE 209 AA; A26AF07CA5D8D5E7 CRC64;				
Query Match 100.0%; Score 1084; DB 2; Length 209;					

```
Best Local Similarity 100.0%; Pred. No. 1.2e-71;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFKLHVNSARQYKDLWNMSDDKFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARN 60
DB 1 MKFKLHVNSARQYKDLWNMSDDKFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARN 60

QY 61 DSVIVADQPTTFLKNCVEEGLFNLASPFENEFKASEDDIKKPLDLSPLATPIIR 120
DB 61 DSVIVADQPTTFLKNCVEEGLFNLASPFENEFKASEDDIKKPLDLSPLATPIIR 120

QY 121 SKIEPVSVEVTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAVRPASIQVNPVLLTSSD 180
DB 121 SKIEPVSVEVTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAVRPASIQVNPVLLTSSD 180

QY 181 SVVITQQAVPSTSTVTITQAPSSNRPIV 209
DB 181 SSVITQQAVPSTSTVTITQAPSSNRPIV 209

RESULT 2
QBN75
ID Q8BN75 PRELIMINARY; PRT; 234 AA.
AC Q8BN75;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
DE library, clone:E130020102 product:activating transcription factor 2,
DE full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20499374; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RX Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RX Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RX Konno H., Akiyama J., Itoh K., Kitsumai T., Tashiro H., Itoh M.,
RX Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RX Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
```

```
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multipillar sequencer.";
RL Genome Res. 10:1757-1771 (2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA Adachi J., Aizawa K., Hanagaki T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Kanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirakawa T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK087424; BAC39869.1; -.
DR HSSP; P15336; 1BHI.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 1.
SQ SEQUENCE 234 AA; 25208 MW; F06750FA9EB37A4D CRC64;
Query Match 90.8%; Score 984; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 2.8e-64;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 MSDDKPFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARNDSVIVADQPTTFLK 78
DB 1 MSDDKPFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARNDSVIVADQPTTFLK 60

QY 79 CEEVGLFNLASPFENEFKASEDDIKKPLDLSPLATPIIRSKIEEPSVVEVTHQDSPL 138
DB 61 CEEVGLFNLASPFENEFKASEDDIKKPLDLSPLATPIIRSKIEEPSVVEVTHQDSPL 120

QY 139 PHPSTTSDEKEVPLAQTAQTSIAVRPASIQVNPVLLTSSDSSVITQQAVPSTSTVI 198
DB 121 PHPSTTSDEKEVPLAQTAQTSIAVRPASIQVNPVLLTSSDSSVITQQAVPSTSTVI 180

QY 199 TOAPSSNRPIV 209
DB 181 TOAPSSNRPIV 191

RESULT 3
QSCBR9
ID Q8CBR9 PRELIMINARY; PRT; 420 AA.
AC Q8CBR9;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Mus musculus adult male urinary bladder cDNA, RIKEN full-length
DE enriched library, clone:9530046122 product:activating transcription
DE factor 2, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 08:05:04 ; Search time 3170 Seconds  
(without alignments)

2509.600 Million cell updates/sec

Title: AAH26175

Perfect score: 1084

Sequence: 1 MKFKLHVNSAQYKDLNMS.....PSTSTVTITQAPSSNRPIV 209

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 69479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/AAH26175/runat 07072005\_101604\_18123/app.query.fasta\_1.391  
-DB=EST -QPMT=fastap -SUFFIX=std.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc.-NORM=ext -HEAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 @CGN 1 1 4352 @runat 07072005\_101604\_18123 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1060	97.8	1606	3 AK087424	AK087424 Mus muscu
2	1056	97.4	904	7 CR763073	CR763073 DKEZp4691
3	1052	97.0	956	5 BU839869	BU839869 AGENCOURT
4	1046.5	96.5	3623	3 AK035438	AK035438 Mus muscu
5	1016	93.7	883	4 B1854668	B1854668 603381603
6	1014	93.5	900	5 BU158697	BU158697 AGENCOURT
7	944	87.1	805	1 AJ454962	AJ454962 AJ454962
8	943.5	87.0	826	1 AJ454852	AJ454852 AJ454852
9	923	85.1	792	6 CA750476	CA750476 UI-M-FY0-

10	920	84.9	744	1	AJ453735
11	918	84.7	802	7	CR629564
12	915.5	84.5	716	2	BF613172
13	915.5	84.5	893	5	EX851498
14	910	83.9	746	6	CD348287
15	910	83.9	749	6	CD348288
16	895	82.6	742	7	CR634629
17	883	81.5	664	7	CF535749
18	883	81.5	782	6	CD348295
19	873.5	80.6	909	2	BF166515
20	866	79.9	670	7	CR367891
21	863	79.6	650	7	CR367893
22	860	79.3	620	5	BP220000
23	859	79.2	686	2	BE882603
24	855.5	78.9	915	5	BQ715489
25	848.5	78.3	781	6	CA512317
26	845	78.0	875	5	BQ891414
27	842	77.7	758	6	CB521135
28	841	77.6	552	9	CL570065
29	840	77.5	757	7	CK635513
30	837	77.2	613	5	BU662818
31	831.5	76.7	2106	3	AK051591
32	826	76.2	908	5	BQ229140
33	810	74.7	806	6	CB961368
34	793	73.2	2835	3	AY724488
35	791.5	73.0	758	6	CA510974
36	791	73.0	924	4	B1551297
37	785	72.4	818	7	CN455466
38	779	71.9	545	2	BE549191
39	779	71.9	577	7	CN367889
40	779	71.9	651	7	CN367890
41	779	71.9	841	5	BU598130
42	773.5	71.4	958	5	BU503620
43	770	71.0	606	5	BQ037883
44	749	69.1	827	4	B1548366
45	744	68.6	722	7	CF532094

#### ALIGNMENTS

RESULT 1

AK087424

LOCUS

DEFINITION

AK087424 1606 bp mRNA linear HTC 03-APR-2004  
Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched  
library, clone: E13020102 product: activating transcription factor  
2, full insert sequence.

ACCESSION

AK087424

VERSION

AK087424.1 GI:26352477

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN Integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
AUTHORS

4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5

The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 1606)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

cdna library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

1. 1606

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:E130020102"

/db\_xref="taxon:10090"

/clone="E130020102"

/tissue\_type="eyeball"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="0 day neonate"

349. 1053

/note="unnamed protein product; activating transcription  
factor 2 (MGD) [MG1:109349, GB|S76655, evidence: BLASTN,  
99%, match=1432]  
putative"

/codon\_start=1

/protein\_id="BAC39869.1"

/db\_xref="GI:26352478"

/translation="MSDDKPFUCTAPGCGQRTNEDHLAVHKKHMTLKTGPARNDS  
VIVADQTPPTPEFLKNCVEVGLFNLASPFENFKASEDDIKKMPDLSPATPIIR  
SKIEPSVETTHQSPLPHPESTTSDEKVLQATQPTSAIVRPASLQPNVLLTS  
SDSVIIQAVPSPTSSTVITQAPSSNRPIVPGVFPULLHLPLNGTMPVAIPASIT  
SSNVHVAAPVPSWKV"

ORIGIN

# Alignment Scores:

Pred. No.: 7,53e-99 Length: 1606  
Score: 1060.00 Matches: 206  
Percent Similarity: 98.56% Conservative: 0  
Best Local Similarity: 98.56% Mismatches: 3  
Query Match: 97.79% Indels: 0  
DB: Gaps: 0

AAH26175 (1-209) x AK087424 (1-1606)

Qy	1	MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTyrAsnMetSer	20
Db	295	ATGAAATTCAGTTGTCATGTAATCTGCCAGATAAAACACGACCTGTGGAATATGAGT	354
Qy	21	AspAspLysProPheLeuCythrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp	40
Db	355	GATGACAAACCTTTCTATGCACTGCCCTGGGTGTGCCACGCGTTTTTACCACAGGAGAT	414
Qy	41	HisLeuAlaValHisLysHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn	60
Db	415	CATTGGCTGTCCATAAACATATAACATGAGATGACACTGAAATTTGGTCCAGCAGTAAAT	474
Qy	61	AspSerValIleValAlaAspGlnThrProThrProThrArgPheLeuLysAsnCyseGlu	80
Db	475	GACAGTGTCAATTGTGGCTGATCAGACTCCAAAGCCCAACACAGATTCTCTAAAAAACTGTGAA	534
Qy	81	GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer	100
Db	535	GAAGTGGGTTTTGTTCAATGAGTTGGCAAGTCCATTTGAAAAATGAATTCAGAAAGGCTTCC	594
Qy	101	GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg	120
Db	595	GAAGATGACATTAATAAATGCTCTAGATTGTGCCCTCTTGGCAACACCATCATAGA	654
Qy	121	SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis	140
Db	655	AGCAAAATTGAGGAGCCTTCTGTTGTGAAACAACTCACACGAGCAGAGCCCTTTACCTCAC	714
Qy	141	ProGluSerThrThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr	160
Db	715	CCCGAGTGCAGTACAGTGATGAAAGAGAGTACCATTTGGCACAACATGCAACGCCACA	774
Qy	161	SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp	180
Db	775	TCAGCTATCGTTTCGTCCAGCATCATTCAGAGTTCCCAATGTGCTGCTCACAGTTCGAC	834
Qy	181	SerSerValIleIleGlnGlnAlaValProSerProThrSerSerThrValIleThrGln	200
Db	835	TCAAGTGTAAATTAATTAACAGCAGTACCTTCCAAACCTCAAGTACTGTATATCACCAG	894
Qy	201	AlaProSerSerAsnArgProIleVal	209
Db	895	GCACCATCTCTAAACAGGCCAAATTTGTT	921

RESULT 2  
CR763073  
LOCUS  
DEFINITION  
DKEZp46911824\_r1 469 (synonym: pkid1) Pongo pygmaeus cdna clone  
DKEZp46911824 5', mRNA sequence.  
CR763073  
ACCESSION  
VERSION  
CR763073.1 GI:52600802  
KEYWORDS  
EST.  
SOURCE  
Pongo pygmaeus (orangutan)  
ORGANISM  
Pongo pygmaeus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.  
REFERENCE  
1 (bases 1 to 904)  
AUTHORS  
Ansong, W., Krieger, S., Regiert, T., Rittmueller, C., Schwager, B.,  
Mewes, H.W., Wei, B., Amid, C., Osanger, A., Fobo, G., Han, M. and  
Wiemann, S.  
TITLE  
Pongo pygmaeus mRNA (Ansong, W., Krieger, S., Regiert, T., et al.)  
JOURNAL  
Unpublished (2004)  
COMMENT  
Contact: MIPS

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: July 8, 2005, 03:11:06 ; Search time 28 Seconds  
(without alignments)  
718.189 Million cell updates/sec

Title: AAH26175  
Perfect score: 1084  
Sequence: 1 MKFKLVNSARQYKDLWNMS.....PSPTSSTVTQAPSSNRPIV 209

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: Pirl:.\*  
2: Pirl2:.\*  
3: Pirl3:.\*  
4: Pirl4:.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1084	100.0	505	1 S05380	transcription fact
2	923	85.1	486	2 JC4028	activating transcr
3	798	73.6	456	2 B42026	cyclic AMP respons
4	719	66.3	448	2 A42026	CAMP response elem
5	711	65.6	389	1 A39429	CAMP response elem
6	561	51.8	483	2 S12741	transcription fact
7	528	48.7	358	2 C42026	cyclic AMP respons
8	390.5	36.0	508	2 A45477	CAMP response elem
9	297	27.4	313	2 A34785	DNA-binding protei
10	133.5	12.3	593	2 JC7829	metal-responsive t
11	115	10.6	565	2 T39863	zinc finger protei
12	110.5	10.2	1367	1 S48478	glucan 1,4-alpha-9
13	110	10.1	447	2 I56511	Zic protein - mous
14	109.5	10.1	428	2 T43532	zinc finger protei
15	109.5	10.1	673	2 S35335	transcription fact
16	106	9.8	1251	2 S49645	probable membrane
17	104.5	9.6	700	1 S01991	transforming prote
18	104	9.6	1046	2 T29776	hypothetical prote
19	104	9.6	1983	2 T00385	KIAA0624 protein -
20	103	9.5	1133	2 S54496	probable membrane
21	102.5	9.5	533	2 JS0304	developmental cont
22	102.5	9.5	644	2 S39356	transcription fact
23	102.5	9.5	672	1 S73715	cytadherence acces
24	102	9.4	1361	2 T30894	neuronal specific DN
25	101	9.3	907	1 Q0BE21	membrane antigen 9
26	100	9.2	474	2 S66480	carbon catabolite
27	100	9.2	686	1 S28050	transforming prote
28	100	9.2	1263	2 T13805	spalt-related prot
29	100	9.2	1350	2 T30341	zinc finger protei

30 99 9.1 394 2 S50486 hypothetical prote  
31 98.5 9.1 457 2 I54340 DNA-binding protei  
32 98 9.0 462 2 T11637 hypothetical prote  
33 97.5 9.0 515 2 A44256 zinc-finger protei  
34 97 8.9 382 2 S71669 finger protein MIG  
35 97 8.9 532 2 C87793 protein C27A12.3  
36 97 8.9 878 2 T21621 hypothetical prote  
37 97 8.9 3375 2 T19821 hypothetical prote  
38 96.5 8.9 1355 2 S40022 spalt protein - fr  
39 96 8.9 609 2 A49839 odd-paired - fruit  
40 96 8.9 753 2 S48059 metal-regulatory t  
41 96 8.9 2109 2 E89065 protein H05O09.1  
42 96 8.9 23247 hypothetical prote  
43 95 8.8 185 2 T49660 glucan 1,4-alpha-9  
44 95 8.8 744 2 T20969 hypothetical prote  
45 95 8.8 1323 2 T30253 spalt protein - mo

ALIGNMENTS

RESULT 1  
S05380  
transcription factor ATF2 - human  
N:Alternate names: activating transcription factor 2 (ATF-2); cAMP response element-bin  
N:Contains: CAMP response element-binding protein HB16  
C:Species: Homo sapiens (man)  
C:Dates: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: S05380; A34776; E34223  
R:Maekawa, T.; Sakura, H.; Kanei-Ishii, C.; Sudo, T.; Yoshimura, T.; Fujisawa, J.I.; Yo:  
EMBO J. 8, 2023-2028, 1989  
A:Title: Leucine zipper structure of the protein CRE-BP1 binding to the cyclic AMP resp  
A:Reference number: S05380; MUID:90005408; PMID:2529117  
A:Accession: S05380  
A:Molecule type: mRNA  
A:Residues: 1-505 <NAE>  
A:Cross-references: UNIPROT:P15336; EMBL:X15875; NID:930214; PIDN:CAA33886.1; PID:93021  
R:Kara, C.J.; Liou, H.C.; Ivashkiv, L.B.; Glimcher, L.H.  
Mol. Cell. Biol. 10, 1347-1357, 1990  
A:Title: A cDNA for a human cyclic AMP response element-binding protein which is distin  
A:Reference number: A34776; MUID:90205810; PMID:2320002  
A:Accession: A34776  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 211-222, 'N', 224-505 <KAR>  
A:Cross-references: GB:M31630; NID:9183787; PIDN:AAA35951.1; PID:9386762  
R:Hai, T.; Liu, F.; Coukos, W.J.; Green, M.R.  
Genes Dev. 3, 2083-2090, 1989  
A:Title: Transcription factor ATF cDNA clones: an extensive family of leucine zipper pr  
A:Reference number: A91622; MUID:90185187; PMID:2516827  
A:Accession: B34223  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 107-357, 'V', 359-465 <HA2>  
C:Genetics:  
A:Gene: GDB:ATF2; CREB2; TREB7; CRE-BP1  
A:Cross-references: GDB:128011; OMIM:123811  
A:Map position: 2q32-2q32  
C:Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homo  
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation  
F:347-387/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 100.0%; Score 1084; DB 1; Length 505;  
Best Local Similarity 100.0%; Pred. No. 5.3e-78;  
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKFKLVNSARQYKDLWNMSDDKPFCLCTAPGCCQRFNEDHLAVHKHHEMTLKFGPARN 60  
Db 1 MKFKLVNSARQYKDLWNMSDDKPFCLCTAPGCCQRFNEDHLAVHKHHEMTLKFGPARN 60  
QY 61 DSVIVADQTPTRFLKNCCEVGLFNELASPFNEFKKASEDDIKOMPLDLSPLATPIIR 120  
Db 61 DSVIVADQTPTRFLKNCCEVGLFNELASPFNEFKKASEDDIKOMPLDLSPLATPIIR 120

QY 121 SKIEPSVETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQVNVLLTSSD 180  
DB 121 SKIEPSVETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQVNVLLTSSD 180  
QY 181 SSVIIQQAVPSPTSSVTITQAPSSNRPIV 209  
DB 181 SSVIIQQAVPSPTSSVTITQAPSSNRPIV 209

RESULT 2  
JC4028  
activating transcription factor 2 - African clawed frog  
N/Alternate names: cyclic AMP-response element-binding protein  
C/Species: Xenopus laevis (African clawed frog)  
C/Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 09-Jul-2004  
C/Accession: JC4028  
R/Villarreal, X.C.; Richter, J.D.  
Gene 153, 225-229, 1995  
A/Title: Analysis of ATF2 gene expression during early Xenopus laevis developme nt.  
A/Reference number: JC4028; MUID:95180723; PMID:7875593  
A/Accession: JC4028  
A/Molecule type: mRNA  
A/Residues: 1-486 <VIL>  
A/Cross-references: UNIPROT:Q91576; GB:U16158; NID:G887779; PIDN:AAA69518.1; PID:G710326  
C/Comment: This protein is a sequence-specific DNA-binding protein that mediates transcr  
C/Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homol  
C/Keywords: leucine zipper; phosphoprotein; transcription regulation  
F/37-398/Region: leucine zipper motif  
F/329-369/Domain: fos/jun DNA-binding domain homology <FJD>  
F/332-364/Region: basic  
F/82/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted  
F/102/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

Query Match 85.1%; Score 923; DB 2; Length 486;  
Best Local Similarity 94.8%; Pred. No. 2.6e-65;  
Matches 182; Conservative 5; Mismatches 3; Indels 2; Gaps 2;

QY 19 MSDDKPFLCTAGCGQRFNEDHLAVHKHKHMTLKFPGARNDSVIVADQPTPTPRFLKN 78  
DB 1 MSDDKPFLCTAGCGQRFNKHDLAVHKHKHMTLKFPGARNDSVIVADQPTPTPRFLKN 60

QY 79 CEEVGLFNLASPFENEFKASEDDIKKMPDLSPPLATPIIRSKIEPSVETTHQDSPL 138  
DB 61 CEEVGLFNLASPFENEFKASEED-KKMPDLSPPLATPIIRSKIEPSVETTHQDSPL 119

QY 139 PHPESTTSDE-KEVPLAQTAQPTSAIVRPASLQVNVLLTSSDSSVIIQQAVPSPTSTV 197  
DB 120 PHPESTSDQHNQVPLAQTAQPTSAIVRPASLQVNVLLTSSDSSVIIQQAVPSPTSTV 179

QY 198 ITQAPSSNRPIV 209  
DB 180 ITQAPSTNSSIV 191

RESULT 3  
B42026  
cyclic AMP response element DNA-binding protein isoform 1 - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Aug-1999  
C/Accession: B42026  
R/Georgopoulos, K.; Morgan, B.A.; Moore, D.D.  
Mol. Cell. Biol. 12, 747-757, 1992  
A/Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activ  
A/Reference number: A42026; MUID:92123199; PMID:1531087  
A/Contents: EL4  
A/Accession: B42026  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-456 <GEO>  
A/Cross-references: GB:S76657; NID:G243428; PIDN:AAB21128.1; PID:G243429  
A/Note: sequence extracted from NCBI backbone (NCBIN:76657, NCBIP:76658)  
C/Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homol

F:297-337/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 73.6%; Score 798; DB 2; Length 456;  
Best Local Similarity 100.0%; Pred. No. 1.7e-55;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 MTLKFGPARNDSVIVADQPTPTPRFLKNCBEVGLFNLASPFENEFKASDDDIKKMPLD 110  
DB 1 MTLKFGPARNDSVIVADQPTPTPRFLKNCBEVGLFNLASPFENEFKASDDDIKKMPLD 60

QY 111 LSPLATPIIRSKIEPSVETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQ 170  
DB 61 LSPLATPIIRSKIEPSVETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQ 120

QY 171 VPNVLLTSSDSSVIIQQAVPSPTSSVTITQAPSSNRPIV 209

DB 121 VPNVLLTSSDSSVIIQQAVPSPTSSVTITQAPSSNRPIV 159

#### RESULT 4

A42026

CAMP response element-binding protein 3 - mouse

C/Species: Mus musculus (house mouse)

C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 26-Aug-1999

C/Accession: A42026

R/Georgopoulos, K.; Morgan, B.A.; Moore, D.D.

Mol. Cell. Biol. 12, 747-757, 1992

A/Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activ

A/Reference number: A42026; MUID:92123199; PMID:1531087

A/Contents: EL4

A/Accession: A42026

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-448 <GEO>

A/Cross-references: GB:S76655; NID:G243426; PIDN:AAB21127.1; PID:G243427

A/Note: sequence extracted from NCBI backbone (NCBIN:76655, NCBIP:76656)

C/Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homol

F:289-329/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 66.3%; Score 719; DB 2; Length 448;

Best Local Similarity 100.0%; Pred. No. 2.9e-49;

Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 DOTPTPRFLKNCBEVGLFNLASPFENEFKASEDDIKKMPDLSPPLATPIIRSKIEEP 126

DB 9 DOTPTPRFLKNCBEVGLFNLASPFENEFKASEDDIKKMPDLSPPLATPIIRSKIEEP 68

QY 127 SWETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQVNVLLTSSDSSVIIQ 186

DB 69 SWETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQVNVLLTSSDSSVIIQ 128

QY 187 QAVPSPTSSVTITQAPSSNRPIV 209

DB 129 QAVPSPTSSVTITQAPSSNRPIV 151

#### RESULT 5

A39429

CAMP response element-binding protein ATF2 - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C/Accession: A39429

R/Kageyama, R.; Sasai, Y.; Nakanishi, S.

J. Biol. Chem. 266, 15525-15531, 1991

A/Title: Molecular characterization of transcription factors that bind to the CAMP respo

A/Reference number: A39429; MUID:91332085; PMID:1714459

A/Accession: A39429

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-389 <KAG>

A/Cross-references: UNIPROT:Q00969; GB:M65148; NID:G206569; PIDN:AAA42013.1; PID:G206570

C/Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homol

C/Keywords: DNA binding; nucleus; transcription regulation

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 10:23:27 ; Search time 593 Seconds  
(without alignments)  
2212.762 Million cell updates/sec

Title: AAH26175  
Perfect score: 1084  
Sequence: 1 MFKLVNSARQYKDLNWS.....PSPTSTVTITQAPSSNRPIV 209

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6330943 seqs, 3139157217 residues

Total number of hits satisfying chosen parameters: 12661886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/AAH26175/runat 07072005 101607 18236/app query.fasta\_1.391  
-DB=Published Applications NA -OFFT=fastap -SUFFIX=std.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=AAH26175 @CNG 1.1 741 @runat 07072005 101607 18236  
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/pct NEW PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/pct NEW PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US10J\_NEW PUB.seq.\*  
23: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW PUB.seq.\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW PUB.seq.\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	569	52.5	2755	20	US-10-357-930-22633	Sequence 22633, A
2	569	52.5	2755	20	US-10-357-930-28476	Sequence 28476, A
3	434	40.0	800	20	US-10-357-930-23141	Sequence 23141, A
4	434	40.0	800	20	US-10-357-930-29011	Sequence 29011, A
5	427	39.4	501	9	US-09-833-790-215	Sequence 215, App
6	390.5	36.0	2637	11	US-09-968-0074A-974	Sequence 974, App
7	390.5	36.0	2637	15	US-10-171-581-285	Sequence 285, App
8	390.5	36.0	2637	21	US-10-843-641A-7444	Sequence 7444, Ap
9	373.5	34.5	2345	20	US-10-723-860-7409	Sequence 7409, Ap
10	332.5	30.7	432	20	US-10-357-930-35258	Sequence 35258, A
11	332.5	30.7	432	20	US-10-357-930-44098	Sequence 44098, A
12	332	30.6	397	20	US-10-357-930-14137	Sequence 14137, A
13	210.5	19.4	509	16	US-10-210-120-82	Sequence 82, Appl
14	210.5	19.4	509	22	US-10-909-035-82	Sequence 82, Appl
15	180.5	16.7	273	20	US-10-357-930-4968	Sequence 4968, Ap
16	116.5	10.7	4168	17	US-10-108-260A-1212	Sequence 1212, Ap
17	116.5	10.7	6682	20	US-10-723-860-5563	Sequence 5563, Ap
18	115.5	10.7	2493	20	US-10-425-115-114107	Sequence 114107,
19	115	10.6	2193	20	US-10-425-115-114963	Sequence 114963,
20	112.5	10.4	3511	20	US-10-723-860-4919	Sequence 4919, Ap
21	112	10.3	3138	9	US-09-880-107-1716	Sequence 1716, Ap
22	112	10.3	3138	17	US-10-172-118-909	Sequence 909, App
23	112	10.3	3138	18	US-10-342-887-909	Sequence 909, App
24	112	10.3	3138	20	US-10-723-860-176	Sequence 176, App
25	111.5	10.3	2976	9	US-10-425-115-47091	Sequence 47091, A
26	110.5	10.2	4104	9	US-09-801-368-107	Sequence 107, App
27	110.5	10.2	4104	19	US-10-793-639-44	Sequence 44, Appl
28	110.5	10.2	5262	16	US-10-240-965-79	Sequence 79, Appl
29	108.5	10.0	1448	9	US-09-864-761-21928	Sequence 21928, A
30	108.5	10.0	1780	17	US-10-108-260A-874	Sequence 874, App
31	108.5	10.0	1947	9	US-09-864-761-4439	Sequence 4439, Ap
32	108.5	10.0	4067	19	US-10-250-889-27	Sequence 27, Appl
33	108	9.9	2979	20	US-10-423-115-113191	Sequence 113191,
34	107.5	9.9	2487	20	US-10-423-115-112880	Sequence 112880,
35	107.5	9.9	5637	20	US-10-425-115-115805	Sequence 115805,
36	107.5	9.9	18717	15	US-10-156-761-927	Sequence 927, App
37	107.5	9.9	30690	19	US-10-204-862A-1	Sequence 1, Appli
38	107.5	9.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
39	107	9.9	2457	17	US-10-305-720-1106	Sequence 1106, Ap
40	106	9.8	1825	15	US-10-244-367-41	Sequence 41, Appl
41	106	9.8	2106	10	US-09-918-624B-57	Sequence 57, Appl
42	105.5	9.7	842	16	US-10-029-386-22926	Sequence 22926, A
43	105.5	9.7	2334	20	US-10-425-115-113926	Sequence 113926,
44	105.5	9.7	2658	20	US-10-425-115-47126	Sequence 47126, A
45	105.5	9.7	3744	19	US-10-115-633-241	Sequence 241, App

#### ALIGNMENTS

##### RESULT 1

US-10-357-930-22633  
; Sequence 22633, Application US/10357930  
; Publication No. US20040259086A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Endege, Wilson  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF  
; HUMAN PROSTATE CANCER  
; TITLE OF INVENTION: HUMAN PROSTATE CANCER  
; FILE REFERENCE: MRI-007BCN  
; CURRENT APPLICATION NUMBER: US/10/357,930  
; CURRENT FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: 09/785,276  
; PRIOR FILING DATE: 2003-02-16

```
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22633
; LENGTH: 2755
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2753..2754, 2755
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-22633

Alignment Scores:
Pred. No.:          3,02e-57          Length:          2755
Score:              569.00           Matches:         126
Percent Similarity: 69.90%           Conservative:    18
Best Local Similarity: 61.17%        Mismatches:      42
Query Match:        52.49%           Indels:          20
DB:                 20               Gaps:            7

AAH26175 (1-209) x US-10-357-930-22633 (1-2755)

QY 11 ArgGlnTyrLysAspLeuTrp-----AsnMetSerAspAspLysProPheLeu 26
Db 65 CGGAGCTACCGCGCTCTCTCTCTATATGAAATATGGGAGACGACAGCGTTTGTG 124

QY 27 CysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAspHisLeuAlaValHisLys 46
Db 125 TGCAATGCCCGCGGTGTGGACAGAGATTACAAACGAGGACCACCTGCGAGTTTCATATA 184

QY 47 HisLysHisGluMetThrLeuLysPheGlyProAlaArgAsnAspSerValIleValala 66
Db 185 CACAAGCATGAGATGACATTGAAATTTGGCCCGAGCCCGAAGTCTCAGTCATCATTCGCA 244

QY 67 AspGlnThrProThrArgPheLeuLysAsnCysGluGluValGlyLeuPheAsn 86
Db 245 GATCAACCGCTACTCCAACTAGATTCTGAGAGACTGTGAGGAGTGGGACTCTTCAAT 304

QY 87 GluLeuAlaSerProPheGluAsnGluPheLysLysAlaSerGluAspAspLysLys 106
Db 305 GAACTAGCTAGCTCTCTTCAACATGAATTCAGAAAGCTGACAGATGAGGATGAGAAAGA 364

QY 107 Met-----ProLeuAspLeuSerProLeuAlaThrProIleArgSerLysIle 123
Db 365 GCTGCTGTGGCGCCCTTGACATGCTCTGCGCTTCCACACGACATCAAAATCAAGAA 424

QY 124 GluGluProSerValValGluThrThrHisGlnAspSerProLeuProHisProGluSer 143
Db 425 GAAGAGCCAGTGGAGTAGATCATCCCGACCTGATAGCCCTGCTAGTCCCTGTTCC 484

QY 144 ThrThrSerAspGluLysGluVal-----ProLeuAlaGlnThrAlaGlnProThr 160
Db 485 CCACCACTGAGGAGAGAGGTTTACCCCAAGACCTGTTCTG---ATCTCTACCCCCACA 541

QY 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuThrSerSerAsp 180
Db 542 CCCACCATGTAGCTCTGGCTCCCTGCT-----CTCCACTGGGCTATGAT 589

QY 181 SerSerValIleIleGlnAlaValProSerProThrSerSerThrValIleThrGln 200
Db 590 CCA-----CTTCATCAACCTTCCCTCCCAACCTCT-----GTCAATCACAG 634
```

```
QY 201 AlaProSerSerAsnArg 206
Db 635 GCTCCACCATCCAAACAGG 652

RESULT 2
US-10-357-930-28476
; Sequence 28476, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28476
; LENGTH: 2755
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2753..2754, 2755
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-28476

Alignment Scores:
Pred. No.:          3,02e-57          Length:          2755
Score:              569.00           Matches:         126
Percent Similarity: 69.90%           Conservative:    18
Best Local Similarity: 61.17%        Mismatches:      42
Query Match:        52.49%           Indels:          20
DB:                 20               Gaps:            7

AAH26175 (1-209) x US-10-357-930-28476 (1-2755)

QY 11 ArgGlnTyrLysAspLeuTrp-----AsnMetSerAspAspLysProPheLeu 26
Db 65 CGGAGCTACCGCGCTCTCTCTATATGAAATATGGGAGACGACAGCGTTTGTG 124

QY 27 CysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAspHisLeuAlaValHisLys 46
Db 125 TGCAATGCCCGCGGTGTGGACAGAGATTACAAACGAGGACCACCTGCGAGTTTCATATA 184

QY 47 HisLysHisGluMetThrLeuLysPheGlyProAlaArgAsnAspSerValIleValala 66
Db 185 CACAAGCATGAGATGACATTGAAATTTGGCCCGAGCCCGAAGTCTCAGTCATCATTCGCA 244

QY 67 AspGlnThrProThrArgPheLeuLysAsnCysGluGluValGlyLeuPheAsn 86
Db 245 GATCAACCGCTACTCCAACTAGATTCTGAGAGACTGTGAGGAGTGGGACTCTTCAAT 304

QY 87 GluLeuAlaSerProPheGluAsnGluPheLysLysAlaSerGluAspAspLysLys 106
Db 305 GAACTAGCTAGCTCTCTTCAACATGAATTCAGAAAGCTGACAGATGAGGATGAGAAAGA 364

QY 107 Met-----ProLeuAspLeuSerProLeuAlaThrProIleArgSerLysIle 123
Db 365 GCTGCTGTGGCGCCCTTGACATGCTCTGCGCTTCCACACGACATCAAAATCAAGAA 424

QY 124 GluGluProSerValValGluThrThrHisGlnAspSerProLeuProHisProGluSer 143
Db 425 GAAGAGCCAGTGGAGTAGATCATCCCGACCTGATAGCCCTGCTAGTCCCTGTTCC 484

QY 144 ThrThrSerAspGluLysGluVal-----ProLeuAlaGlnThrAlaGlnProThr 160
Db 485 CCACCACTGAGGAGAGAGGTTTACCCCAAGACCTGTTCTG---ATCTCTACCCCCACA 541

QY 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuThrSerSerAsp 180
Db 542 CCCACCATGTAGCTCTGGCTCCCTGCT-----CTCCACTGGGCTATGAT 589

QY 181 SerSerValIleIleGlnAlaValProSerProThrSerSerThrValIleThrGln 200
Db 590 CCA-----CTTCATCAACCTTCCCTCCCAACCTCT-----GTCAATCACAG 634
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 08:11:07 ; Search time 183 Seconds  
(without alignments)

1868.753 Million cell updates/sec

Title: AAH26175

Perfect score: 1084

Sequence: 1 MKFKLVNSARQYKDLWNMS.....PSPTSSTVITQAPSSNRPIV 209

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=std.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosun62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 @CGN 1 1 177 @runat 07072005 101605 18139 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1084	100.0	1621	4	US-09-949-016-5410
2	407.5	37.6	65966	4	US-09-949-016-17152
3	263	24.3	229	4	US-09-513-999C-2946
C	148	13.7	601	4	US-09-949-016-189073
C	143	13.2	601	4	US-09-949-016-189064
5	112	10.3	3138	3	US-09-234-332-5
6	110.5	10.2	4104	4	US-09-614-221A-44
7	108.5	10.0	2394	4	US-09-949-016-5520
8	108.5	10.0	2428	4	US-09-949-016-17262
9	107.9	9.9	2457	4	US-09-016-434-1106
10	107	9.9	2457	4	US-09-016-434-1106
11	106	9.8	1825	4	US-09-342-325C-41
12	104.5	9.6	2627	4	US-09-023-655-1251

13	104.5	9.6	2638	1	US-08-306-691B-46	Sequence 46, Appl
14	104.5	9.6	2653	4	US-09-976-594-66	Sequence 66, Appl
15	104	9.6	6745	4	US-09-774-528-286	Sequence 286, Appl
16	102.5	9.5	3068	2	US-08-224-482-1	Sequence 1, Appl
17	102.5	9.5	3086	1	US-08-040-548-15	Sequence 15, Appl
18	102.5	9.5	3086	1	US-08-466-344-15	Sequence 15, Appl
19	102.5	9.5	3086	6	5206152-1	Patent No. 5206152
20	102.5	9.5	3086	6	5206152-1	Patent No. 5206152
21	102	9.4	589	4	US-09-229-037-29	Sequence 29, Appl
22	102	9.4	589	4	US-09-478-681-29	Sequence 29, Appl
23	102	9.4	589	4	US-09-779-233-17	Sequence 17, Appl
24	102	9.4	589	4	US-09-706-243A-29	Sequence 29, Appl
25	102	9.4	2447	4	US-09-949-016-5274	Sequence 5274, Ap
26	101.5	9.4	3240	4	US-09-949-016-5548	Sequence 5548, Ap
27	101.5	9.4	11929	4	US-09-949-016-17290	Sequence 17290, A
28	101	9.3	3833	1	US-08-917-320-18	Sequence 18, Appl
29	101	9.3	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
30	101	9.3	5931	3	US-08-783-774-1	Sequence 1, Appl
31	101	9.3	5931	3	US-09-556-706B-1	Sequence 1, Appl
32	101	9.3	5931	4	US-09-724-418A-1	Sequence 1, Appl
33	100.5	9.3	2901	4	US-09-342-325C-43	Sequence 43, Appl
34	100	9.2	292	4	US-09-395-448-14	Sequence 14, Appl
35	100	9.2	292	4	US-09-325-796-14	Sequence 14, Appl
36	100	9.2	292	4	US-09-941-450-14	Sequence 14, Appl
37	100	9.2	4674	4	US-09-410-551B-26	Sequence 26, Appl
38	100	9.2	4674	4	US-09-940-316B-26	Sequence 26, Appl
39	99.5	9.2	918	4	US-10-029-180-59	Sequence 59, Appl
40	99.5	9.2	136058	4	US-09-949-016-12565	Sequence 12565, A
41	99.5	9.2	136480	4	US-09-949-016-17064	Sequence 17064, A
42	99	9.1	7785	4	US-09-573-080A-32	Sequence 32, Appl
43	99	9.1	24735	4	US-09-949-016-12850	Sequence 12850, A
44	98.5	9.1	2655	4	US-09-799-451-187	Sequence 187, App
45	97	8.9	3020	4	US-09-220-132-19	Sequence 19, Appl

#### ALIGNMENTS

##### RESULT 1

; Sequence 5410, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE OF INVENTION: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5410  
; LENGTH: 1621  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-5410

Alignment Scores:  
Pred. No.: 8.96e-121 Length: 1621  
Score: 1084.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

AAH26175 (1-209) x US-09-949-016-5410 (1-1621)

Oy 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20

```
Db 24 ATGAATTCAGTTACATGTGAATCTGCCAGCAATACAGACCTGTGGATATGAGT 83
Qy 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40
Db 84 GATGACAAACCCCTTCTATGTACTGCGCTGATGTGGCCAGGGTTTACCAACGAGGAT 143
Qy 41 HisLeuAlaValHisIlyshisIlyshisGluMetThrLeuLysPheGlyProAlaArgAsn 60
Db 144 CATTTGGCTGTCCATAAACAATACATGAGATGACACTGAAATTTGGTCAGCAGCTAAT 203
Qy 61 AspSerValIleValAlaAspGlnThrProThrProThrArgPheLeuLysAsnCysGlu 80
Db 204 GACAGTGTCTTGTGGCTGATCAGACCCCAACACCAACCAAGATTCTTGAACAACTGTGAA 263
Qy 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100
Db 264 GAAGTGGGTTTGTAAATGAGTTGGCGAGTCCATTGTGAAATGAATTCAGAAAGCTTCA 323
Qy 101 GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120
Db 324 GAAGATGACATTAATAAATGCTCTAGATTTATCCCTCTTGCACACCTATCATAGA 383
Qy 121 SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140
Db 384 AGCAAAATTTAGAGAGCTCTGTGTGAGAAACAACCTCAGAGATAGTCTTTACTCTAC 443
Qy 141 ProGluSerThrThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr 160
Db 444 CCAGAGTCTACTACCACTGATGAGAAGGAAGTACCAATTGGCACAACTGCACAGCCACA 503
Qy 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180
Db 504 TCAGCTATTGTTCGTCAGCATCATACAGGTTCCCAATGTGCTGTACAAAGTTCTGAC 563
Qy 181 SerSerValIleIleGlnGlnAlaValProSerProThrProThrSerThrValIleThrGln 200
Db 564 TCAAGTGTATTATTTCAGCAGGAGTACCTTCACCAACCTCAAGTACTGTAATCAACCCAG 623
Qy 201 AlaProSerSerAsnArgProIleVal 209
Db 624 GCACCATCTCTTAACAGGCCAATTGTC 650
```

## RESULT 2

```
US-09-949-016-17152
; Sequence 17152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17152
; LENGTH: 65966
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17152
```

Alignment Scores: 1.89e-36 Length: 65966

Pred. No.: 407.50 Matches: 86

Percent Similarity: 61.38% Conservative: 3

Best Local Similarity: 59.31% Mismatches: 3

```
Query Match: 37.59% Indels: 53
DB: 4 Gaps: 2
AAH26175 (1-209) x US-09-949-016-17152 (1-65966)
Qy 49 HisGluMetThrLeuLysPheGlyProAlaArgAsnAspSerValIleValAlaAspGln 68
Db 20072 CATGATTACTTCTTGCACATTT-----ATTTCAGATCAG 20104
Qy 69 ThrProThrProThrArgPheLeuLysAsnCysGluValGlyLeuPheAsnGluLeu 88
Db 20105 ACCCCACACCAACAAGATTCTTGAACAACTGTGAAGAAGTGGGTTTCTTTAATCAGTTG 20164
Qy 89 AlaSerProPheGluAsnGluPheLysLysAlaSerGluAspAspIleLysLys----- 106
Db 20165 GCGAGTCCATTGTGAGAAATTCAGAAAGCTTCAGAAAGATGACATTTAAAAAAGATGTT 20224
Qy 106 ----- 106
Db 20225 TGTAACCATGAATTTAAAAAGAGTACTTCTGCACAGAAAGTAATCATAGTTTATCTTC 20284
Qy 106 ----- 106
Db 20285 CTGCTTTTAAATTTAGACACACTGTTTGTGAACCTTATAACCTTTCTTATTATTGAA 20344
Qy 107 -----MetProLeuAspLeuSerProLeuAlaThrProIleIleArgSerLysIleGlu 124
Db 20345 TTATAGATGCTCTAGATTTATCCCTCTTGCACACCTATCATAGAAGCAAAATTGAG 20404
Qy 125 GluProSerValValGluThrThrHisGlnAspSerProLeuProHisProGluSerThr 144
Db 20405 GAGCTTCTGTGTGAGAAACAACCTCACCAGGATAGTCTTTTACCTCACCCAGAGTCTACT 20464
Qy 145 ThrSerAspGluLys 149
Db 20465 ACCAGTGATGAGAAG 20479
RESULT 3
US-09-513-999C-2946
; Sequence 2946, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 2946
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 72..227
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 87
; OTHER INFORMATION: y=c or t
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 6
; OTHER INFORMATION: Xaa=Pro or Ser
US-09-513-999C-2946
Alignment Scores: 6.4e-23 Length: 229
Pred. No.: 6.4e-23
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 03:47:39 ; Search time 530 Seconds  
(without alignments)  
2334.389 Million cell updates/sec

Title: AAH26175  
Perfect score: 1084  
Sequence: 1 MKFKLVHNSARQYKDLWNMS.....PSPTSTVTITQAPSSNRPIV 209

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2.1/USPTO.spool/AAH26175/runat.07072005.101602.18107/app\_query.fasta.1.391  
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUFFIX=std.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 -CGEN 1.1 644 @runat.07072005.101602.18107 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_16Dec04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004as: \*  
13: Geneseqn2004bs: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1084	100.0	1647	10	Adk67038 Gene #128
2	1084	100.0	1647	10	Adk66978 Gene #68
3	1062	98.0	1852	10	Adbs3173 Primary r
4	940.5	86.8	3507	13	Adr08447 Full leng
5	571	52.7	1239	12	Adp80924 Mouse C13

6	569	52.5	2755	5	ABV28463	Abv28463 Human pro
7	569	52.5	2755	5	ABV22640	Abv22640 Human pro
8	566	52.2	3217	6	AB199202	Ab199202 Mouse isc
9	434	40.0	800	5	ABV28993	Abv28993 Human pro
10	434	40.0	800	5	ABV23152	Abv23152 Human pro
11	427	39.4	501	6	AAS61674	Aas61674 Lung smal
12	390.5	36.0	2319	10	ADC18732	Adc18732 DNA of hu
13	390.5	36.0	2637	6	ABL69107	Ab169107 Kidney ca
14	390.5	36.0	2637	6	ABK83799	Abk83799 Human cdn
15	390.5	36.0	2637	10	ADH29017	Adh29017 Human chr
16	373.5	34.5	2345	12	ADQ24589	Adq24589 Human sof
17	332.5	30.7	432	5	ABV44079	Abv44079 Human pro
18	332.5	30.7	432	5	ABV35240	Abv35240 Human pro
19	332	30.6	397	5	ABV14146	Abv14146 Human pro
20	263	24.3	229	3	AAC02948	Aac02948 Human sec
21	210.5	19.4	509	10	ADD18510	Ad18510 Human pro
22	180.5	16.7	273	5	ABV04977	Abv04977 Human pro
23	174	16.1	1080	4	ABL21331	Ab121331 Drosophil
24	170.5	15.7	3199	4	ABL21330	Ab121330 Drosophil
25	116.5	10.7	4168	11	ADM02527	Adm02527 Human cdn
26	116.5	10.7	6682	12	ADQ22743	Adq22743 Human sof
27	112.5	10.4	3511	12	ADQ22099	Adq22099 Human sof
28	112	10.3	3138	3	AA62683	Aas62683 Human Zic
29	112	10.3	3138	6	ADN95218	Adn95218 Gene #171
30	112	10.3	3138	12	ADN04235	Adn04235 Antipsori
31	112	10.3	3138	12	ADQ17359	Adq17359 Human sof
32	112	10.3	3138	13	ADR25048	Adr25048 Breast ca
33	111.5	10.3	2947	6	AB199239	Ab199239 Mouse isc
34	110.5	10.2	4434	5	AB199239	Ab199239 DNA encod
35	110.5	10.2	5262	6	AAS79667	Aas79667 Human DNA
36	109.5	10.1	12381	3	AAS94824	Aas94824 Scroptomy
37	109	10.1	2429	4	ABL06615	Ab106615 Drosophil
38	109	10.1	9571	4	ABL06614	Ab106614 Drosophil
39	108.5	10.0	1448	4	AAI24539	Aai24539 Probe #14
40	108.5	10.0	1448	4	ABA69703	Abag69703 Human foe
41	108.5	10.0	1448	4	AAI49785	Aai49785 Probe #18
42	108.5	10.0	1448	4	ABA51616	Abas51616 Human bre
43	108.5	10.0	1448	4	ABA36608	Abak36608 Probe #15
44	108.5	10.0	1448	4	AAK43762	Aak43762 Human bon
45	108.5	10.0	1448	4	AAK17891	Aak17891 Human bra

ALIGNMENTS

RESULT 1  
ID ADK67038 standard; DNA; 1647 BP.  
XX  
AC ADK67038;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Gene #128 for inhibitory RNA to manipulate stem cell phenotype.  
XX  
KW ds; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;  
XX  
KW pluripotent stem cell.  
XX  
OS Homo sapiens.  
XX  
PN WO2003068961-A2.  
XX  
PD 21-AUG-2003.  
XX  
PF 12-FEB-2003; 2003WO-GE0000579.  
XX  
PR 13-FEB-2002; 2002GB-00003359.  
XX  
PR 13-FEB-2002; 2002GB-00003387.  
XX  
PA (AXOR-) AXORDIA LTD.  
XX  
PI Andrews P, Walsh J, Gokhale P;  
XX  
DR WPI; 2003-697528/66.

```
XX New inhibitory RNA molecule having double stranded RNA molecules, useful
PT for manipulating the phenotype of stem cells, preferably pluripotent
PT stem cells.
XX
XX Disclosure; SEQ ID NO 128; 157pp; English.
XX
XX The invention relates to an inhibitory RNA (RNAi) molecule derived from a
CC nucleic acid molecule comprising a defined nucleic acid sequences given
CC in the specification or a sequence which hybridizes to the sequences and
CC encodes a Notch signaling target gene or which is a degenerate as a
CC result of the genetic code of the sequences. The methods and compositions
CC of the present invention are useful for manipulating the phenotype of
CC stem cells, preferably pluripotent stem cells. This sequence corresponds
CC to one of the nucleic acid molecules of the invention.
XX
XX Sequence 1647 BP; 504 A; 396 C; 341 G; 406 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 3.75e-97 Length: 1647
Score: 1084.00 Matches: 209
Percent Similarity: 100.00% Conservatative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
AAH26175 (1-209) x ADK67038 (1-1647)
QY 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20
DB 27 ATGAAATTCAGTTACATGTGAATTCGCCAGCAATACAGGACCTGTGGATATGAGT 86
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40
DB 87 GATGACAAACCCCTTCTATGTACTCGCTGATGCGCCAGCTTTTACCAACGAGGAT 146
QY 41 HisLeuAlaValHisLysHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60
DB 147 CATTTGGCTGTCTATAAACAATAAACATGAGATGACACTGAAATTTGGTCCAGCAGTAA 206
QY 61 AspSerValIleValAlaAspGlnThrProThrProThrArgPheLeuLysAsnGlu 80
DB 207 GACAGTGTCATTGTGGCTGATCAGACCCCAACACCAACAGATTCTTGAAAACTGTGA 266
QY 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100
DB 267 GAAGTGGGTTCTTTAATGAGTTGGCGAGTCCATTGAGAAATGAATTCAGAAAGCTTCA 326
QY 101 GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120
DB 327 GAAGATGACATTAATAAATAATGCTCTAGATTATCCCTCTTGCACACCTATCATAGA 386
QY 121 SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140
DB 387 ACCAAAATTGAGAGCCCTCTCTGTGTAGAAACAACCTCACCAGATAGTCTTTACTCTAC 446
QY 141 ProGluSerThrThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr 160
DB 447 CCAGAGTCTACTACCACTGATGAGAAGGAAGTACCATTGGCACAACATGCACAGCCACA 506
QY 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180
DB 507 TCAGCTATTGTTCGTCACGATCATACAGGTTCCTCAATGCTGCTCTACAGTTCTGAC 566
QY 181 SerSerValIleIleGlnGlnAlaValProSerProThrSerSerThrValIleThrGln 200
DB 567 TCAAGTGTAATTAATTCAGCAGGAGTACCTTCAACCAACCTCAAGTACTGTAACTACCCAG 626
QY 201 AlaProSerSerAsnArgProIleVal 209
DB 627 GCACCATCTCTTAACAGGCGCAATTGTG 653
```

RESULT 2

```
ADK66978
ID ADK66978 standard; DNA; 1647 BP.
XX
XX AC ADK66978;
XX
XX 06-MAY-2004 (first entry)
XX
XX Gene #68 for inhibitory RNA to manipulate stem cell phenotype.
XX
XX ds; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;
XX pluripotent stem cell.
XX
XX Homo sapiens.
XX
XX WO2003068961-A2.
XX
XX 21-AUG-2003.
XX
XX 12-FEB-2003; 2003WO-GB000579.
XX
XX 13-FEB-2002; 2002GB-00003359.
XX
XX 13-FEB-2002; 2002GB-00003387.
XX
XX (AXOR-) AXORDIA LTD.
XX
XX Andrews P, Walsh J, Gokhale P;
XX
XX WPI; 2003-697528/66.
XX
XX New inhibitory RNA molecule having double stranded RNA molecules, useful
XX for manipulating the phenotype of stem cells, preferably pluripotent
XX stem cells.
XX
XX Disclosure; SEQ ID NO 68; 157pp; English.
XX
XX The invention relates to an inhibitory RNA (RNAi) molecule derived from a
XX nucleic acid molecule comprising a defined nucleic acid sequences given
XX in the specification or a sequence which hybridizes to the sequences and
XX encodes a Notch signaling target gene or which is a degenerate as a
XX result of the genetic code of the sequences. The methods and compositions
XX of the present invention are useful for manipulating the phenotype of
XX stem cells, preferably pluripotent stem cells. This sequence corresponds
XX to one of the nucleic acid molecules of the invention.
XX
XX Sequence 1647 BP; 504 A; 396 C; 341 G; 406 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 3.75e-97 Length: 1647
Score: 1084.00 Matches: 209
Percent Similarity: 100.00% Conservatative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
AAH26175 (1-209) x ADK66978 (1-1647)
QY 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20
DB 27 ATGAAATTCAGTTACATGTGAATTCGCCAGCAATACAGGACCTGTGGATATGAGT 86
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40
DB 87 GATGACAAACCCCTTCTATGTACTCGCTGATGCGCCAGCTTTTACCAACGAGGAT 146
QY 41 HisLeuAlaValHisLysHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60
DB 147 CATTTGGCTGTCTATAAACAATAAACATGAGATGACACTGAAATTTGGTCCAGCAGTAA 206
QY 61 AspSerValIleValAlaAspGlnThrProThrProThrArgPheLeuLysAsnGlu 80
DB 207 GACAGTGTCATTGTGGCTGATCAGACCCCAACACCAACAGATTCTTGAAAACTGTGA 266
QY 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 06:13:59 ; Search time 4142 Seconds

(without alignments)  
2444.988 Million cell updates/sec

Title: AAH26175

Perfect score: 1084

Sequence: 1 MKFKLVNSARQYKDLNWS.....PSTSTVTITQAPSSNRPIV 209

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 segs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh  
-Q=/cgn2.1/USPTO.spool/AAH26175/runat.07072005.101603.18114/app.query.fasta.1.391  
-DB=GenEmbl -QFMT=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAFSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=AAH26175 @CGN 1 1 4200 @runat.07072005.101603.18114 -NCPUS=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_em.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1084	100.0	1370	9 BC026175	BC026175 Homo sapi
2	1084	100.0	1647	6 AX821940	AX821940 Sequence
3	1084	100.0	1647	6 AX822000	AX822000 Sequence
4	1084	100.0	1647	9 HSCREBP1	X15875 Human mRNA

5	1062	98.0	1852	10 RNU38938	U38938 Rattus norv
6	1013.5	93.5	3957	10 BC042210	BC042210 Mus muscu
7	1001	92.3	1812	5 GGA17724	Y17724 Gallus gall
8	984	90.8	1464	10 AF483482	AF483482 Mus muscu
9	984	90.8	1464	10 AF483483	AF483483 Mus muscu
10	949.5	87.6	1596	10 S76657	S76657 cyclic AMP
11	940.5	86.8	3507	6 CQ851484	CQ851484 Sequence
12	940.5	86.8	3507	9 AK128731	AK128731 Homo sapi
13	937	86.4	1520	5 XLU16158	U16158 Xenopus lae
14	869	80.2	3983	9 HSM808617	BM648469 Homo sapi
15	845	78.0	1848	10 S76655	S76655 cyclic AMP
16	831.5	76.7	4130	10 BC079883	BC079883 Mus muscu
17	790	72.9	3909	10 BC082596	BC082596 Mus muscu
18	746	68.8	1209	10 RATRATF2	M65148 Rat Rattf2 m
19	679.5	62.7	1302	10 S76659	S76659 cyclic AMP
20	578	53.3	2867	10 BC026483	BC026483 Mus muscu
21	569	52.5	2755	6 CQ490766	CQ490766 Sequence
22	569	52.5	2755	6 CQ496609	CQ496609 Sequence
23	569	52.5	2758	9 HSATFA	X52943 Human mRNA
24	566	52.2	3217	6 AX305250	AX305250 Sequence
25	566	52.2	3217	10 MUSTCRATF	M77167 Mouse T-cell
26	555.5	51.2	1485	9 HSATFA1	X57197 H.sapiens m
27	533	49.2	4851	12 AF050498	AF050498 Fusion tr
28	504.5	46.5	1227	6 CQ720976	CQ720976 Sequence
29	434	40.0	800	6 CQ491274	CQ491274 Sequence
30	434	40.0	800	6 CQ497144	CQ497144 Sequence
31	427	39.4	501	6 AX321198	AX321198 Sequence
32	423	39.0	787	9 BC042363	BC042363 Homo sapi
33	409.5	37.8	725	5 BX933040	BX933040 Gallus ga
34	409.5	37.8	867	5 BX930883	BX930883 Gallus ga
35	407.5	37.6	176033	2 ALJ161794	ALJ161794 Homo sapi
36	407.5	37.6	178342	9 AC007435	AC007435 Homo sapi
37	398.5	36.8	461	6 CQ727434	CQ727434 Sequence
38	390.5	36.0	2319	9 HUMCRBPAB	L05911 Human cAMP
39	390.5	36.0	2837	6 AX336935	AX336935 Sequence
40	390.5	36.0	2837	9 HUMCRBPAB	L05515 Homo sapien
41	387.5	35.7	173793	2 AC129355	AC129355 Rattus no
42	386	35.6	207639	10 AL844581	AL844581 Mouse DNA
43	366.5	33.8	166766	2 AC119682	AC119682 Rattus no
44	332.5	30.7	432	6 CQ503391	CQ503391 Sequence
45	332.5	30.7	432	6 CQ512231	CQ512231 Sequence

## ALIGNMENTS

RESULT 1	BC026175	1370 bp	mRNA	linear	PRI 16-SEP-2003
LOCUS	Homo sapiens activating transcription factor 2, mRNA (cDNA clone IMAGE:4308091), complete cds.				
DEFINITION	Homo sapiens activating transcription factor 2, mRNA (cDNA clone IMAGE:4308091), complete cds.				
ACCESSION	BC026175				
VERSION	BC026175.1 GI:20072896				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1370) Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altshuler,S.F., Zeeberg,B., Buetow,K.H., Schafer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heideh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullah,J.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Heiton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,				

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257  
12477932  
2 (bases 1 to 1370)  
Direct Submission  
Strausberg, R.  
Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgi.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: ATCC/DC/DT/PTP  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amadan@systemsbio.org](mailto:amadan@systemsbio.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 38 Row: c Column: 20  
This clone has the following problem: The cds is short compared to the longest cds in the locus.

## FEATURES

source

1..1370  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4308091"  
/tissue\_type="Skin, melanotic melanoma, high MDR."  
/clone\_lib="NIH MGC\_62"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
1..1370  
/gene="ATF2"  
/note="synonyms: CRE-BP1, TREB7, HB16"  
/db\_xref="LocusID:1386"  
/db\_xref="MIM:123811"  
267..896  
/gene="ATF2"  
/codon\_start=1  
/product="ATF2 protein"  
/protein\_id="AAH26175.1"  
/db\_xref="GI:20072897"  
/db\_xref="LocusID:1386"  
/db\_xref="MIM:123811"  
/translation="MKFKLVNSRQYKDLWNMSDDKFLCTAPCCGQRFNEDHLAV HKHKHEMTLKGFPANESVIVADQPTTFLKNCSEVGLFNELASPFENPFKASED DIKMPDLSPATPIIRSKIEEPSVETTHQDSPSPSTTSDEKVEPLAQAPT SAIVRPASIQVNPVLLTSSDSVITQAVPSPTSTVITQAPSSNRPIV"

## ORIGIN

Alignment Scores:

Pred. No.: 8,568-76 Length: 1370  
Score: 1084.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

AAH26175 (1-209) x BC026175 (1-1370)

Qy 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTyrAsnMetSer 20

Db 267 ATGAATTCAGTTACATGTGAATTCGCCAGGCAATACAGGACCTGTGGAATATGAGT 326  
Qy 21 AspAspLysProPheLeuCythrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40  
Db 327 GATGACAAACCTTCTATGTACTGCGCTGGATGTGGCCAGCGTTTTACCAACGAGGAT 386  
Qy 41 HisLeuAlaValHisLysHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60  
Db 387 CATTTGGCTGTCCATAAACAATGAGATGACACTGAAATTTGGTCCAGCAGCTAAT 446  
Qy 61 AspSerValIleValAlaAspGlnThrProThrArgPheLeuLysAsnGlu 80  
Db 447 GACAGTGTCAATGGCTGATCAGACCCCAACCAACAGATTTCTTGAAAACTGTGAA 506  
Qy 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100  
Db 507 GAAGTGGGTTTGTATTAATGAGTGGCGAGTCCATTTGAGAATGAATTCAGAAAGCTTCA 566  
Qy 101 GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120  
Db 567 GAAGATGACATTAATAAATGCTCTAGATTTATCCCTCTTGCAACACCTATCATAGA 626  
Qy 121 SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140  
Db 627 AGCAAAATTGAGGAGCCTTCTGTGTAGAAACAACACTCACAGGATAGTCTTACTCTAC 686  
Qy 141 ProGluSerThrThrSerAspGluGluValProLeuAlaGlnThrAlaGlnProThr 160  
Db 687 CCAGAGTCTACTACAGTGTAGAGAGAGTACCATTTGGCACAACATGCACAGGCCACA 746  
Qy 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180  
Db 747 TCAGCTATTGTTGCTCCAGCATCATACAGTTCCCAATGTGCTTACAGTTCTGCAC 806  
Qy 181 SerSerValIleIleGlnAlaValProSerProThrSerSerThrValIleThrGln 200  
Db 807 TCAAGTGTAAATATTACAGCAGCAGTACCTTCACCACTCAAGTACTGTATACCCAG 866  
Qy 201 AlaProSerSerAsnArgProIleVal 209  
Db 867 GCACCATCTCTCTAACAGGCCAATTGTG 893

## RESULT 2

AX821940 1647 bp DNA linear PAT 10-DEC-2003  
LOCUS Sequence 68 from Patent WO03068961.  
DEFINITION AX821940  
ACCESSION AX821940  
VERSION AX821940.1 GI:39725161

## KEYWORDS

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 Andrews, P.A., Walsh, J.A. and Gokhale, P.A.  
Method to modify differentiation of pluripotential stem cells  
Patent: WO 03068961-A 68 21-AUG-2003;  
Axordia Limited (GB)

## FEATURES

Location/Qualifiers

source

1..1647  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:

Pred. No.: 1,058-75 Length: 1647  
Score: 1084.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 03:17:01 ; Search time 90 Seconds  
(without alignments)  
897.145 Million cell updates/sec

Title: Aah26175  
Perfect score: 1084  
Sequence: 1 MFKLVNSARQYKDLNMS.....PSTSTVTITQAPSSNRPIV 209

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1726216 seqs, 386330316 residues

Total number of hits satisfying chosen parameters: 1726216

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pcp.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pcp.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pcp.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pcp.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pcp.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pcp.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pcp.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pcp.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pcp.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	14.4	29	9 US-09-970-515-20	Sequence 20, Appl
2	156	14.4	29	14 US-10-165-250A-20	Sequence 20, Appl
3	156	14.4	29	15 US-10-342-683-20	Sequence 20, Appl
4	156	14.4	29	15 US-10-457-614A-20	Sequence 20, Appl
5	156	14.4	29	17 US-10-924-028-20	Sequence 20, Appl
6	144	13.7	27	13 US-10-115-178-6	Sequence 6, Appl
7	115.5	10.7	830	16 US-10-425-115-298770	Sequence 298770,
8	112	10.3	447	16 US-10-723-860-177	Sequence 177, App
9	111.5	10.3	991	16 US-10-425-115-231754	Sequence 231754,
10	110.5	10.2	1292	17 US-10-741-600-1572	Sequence 1572, Ap
11	110.5	10.2	1367	9 US-09-801-368-108	Sequence 108, App

12	110	10.1	730	16	US-10-425-115-299626	Sequence 299626,
13	108	10.0	992	16	US-10-425-115-297854	Sequence 297854,
14	106	9.8	443	14	US-10-244-367-42	Sequence 42, Appl
15	105.5	9.7	885	16	US-10-425-115-231789	Sequence 231789,
16	105	9.7	960	16	US-10-425-115-298810	Sequence 298810,
17	105	9.7	3394	16	US-10-408-765A-529	Sequence 529, App
18	104.5	9.6	700	14	US-10-171-311-160	Sequence 160, App
19	104.5	9.6	700	14	US-10-301-822-141	Sequence 141, App
20	104.5	9.6	700	16	US-10-751-736-103	Sequence 103, Appl
21	104.5	9.6	700	17	US-10-645-756-36	Sequence 36, Appl
22	104.5	9.6	777	16	US-10-425-115-298589	Sequence 298589,
23	104	9.6	894	16	US-10-425-115-231793	Sequence 231793,
24	104	9.6	1983	16	US-10-723-860-1890	Sequence 1890, Ap
25	103	9.5	492	16	US-10-425-115-298835	Sequence 298835,
26	103	9.5	1166	15	US-10-104-047-2949	Sequence 2949, Ap
27	102.5	9.5	533	15	US-10-373-238-17	Sequence 17, Appl
28	102	9.4	1689	16	US-10-425-115-300468	Sequence 300468,
29	102	9.4	1845	17	US-10-732-923-8645	Sequence 8645, Ap
30	101.5	9.4	619	16	US-10-739-930-3129	Sequence 9129, Ap
31	101.5	9.4	999	16	US-10-425-115-231720	Sequence 231720,
32	101	9.3	533	9	US-09-833-790-428	Sequence 428, App
33	101	9.3	878	16	US-10-722-050-2	Sequence 2, Appli
34	101	9.3	907	9	US-09-328-599A-1	Sequence 1, Appli
35	101	9.3	910	16	US-10-425-115-299670	Sequence 299670,
36	101	9.3	501	14	US-10-244-367-44	Sequence 44, Appl
37	100.5	9.3	1087	16	US-10-425-115-300675	Sequence 300675,
38	100.5	9.2	97	9	US-09-925-796-15	Sequence 15, Appl
39	100	9.2	97	9	US-09-941-450-15	Sequence 15, Appl
40	100	9.2	97	9	US-09-942-090-15	Sequence 15, Appl
41	100	9.2	97	16	US-10-843-944-15	Sequence 15, Appl
42	100	9.2	97	17	US-10-922-546-15	Sequence 15, Appl
43	100	9.2	267	9	US-09-864-761-37826	Sequence 37826, A
44	100	9.2	1557	10	US-09-940-316B-27	Sequence 27, Appl
45	100	9.2				

ALIGNMENTS

RESULT 1  
US-09-970-515-20  
; Sequence 20, Application US/09970515  
; Patent No. US20020127676A1  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501 DIV  
; CURRENT APPLICATION NUMBER: US/09/970,515  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: USSN 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-09-970-515-20

Query Match 14.4%; Score 156; DB 9; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 37 TNEHLAVHKHKHMTLKFGPARNDSVIV 65  
Db 1 TNEHLAVHKHKHMTLKFGPARNDSVIV 29

RESULT 2  
US-10-165-250A-20  
; Sequence 20, Application US/10165250A  
; Publication No. US20030108539A1  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501 CIP  
; CURRENT APPLICATION NUMBER: US/10/165,250A  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: USSN 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-10-165-250A-20

Query Match 14.4%; Score 156; DB 14; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 37 TNEHDLAVHKHKHEMTLKFGRNDSVIV 65  
|||||  
Db 1 TNEHDLAVHKHKHEMTLKFGRNDSVIV 29

RESULT 3  
US-10-342-683-20  
; Sequence 20, Application US/10342683  
; Publication No. US20030220480A1  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501  
; CURRENT APPLICATION NUMBER: US/10/342,683  
; PRIOR FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: US/09/503,954A  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: USSN 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-10-342-683-20

Query Match 14.4%; Score 156; DB 15; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 37 TNEHDLAVHKHKHEMTLKFGRNDSVIV 65  
|||||  
Db 1 TNEHDLAVHKHKHEMTLKFGRNDSVIV 29

RESULT 4  
US-10-457-614A-20  
; Sequence 20, Application US/10457614A  
; Publication No. US20040082509A1  
; GENERAL INFORMATION:

; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501CIP2  
; CURRENT APPLICATION NUMBER: US/10/457,614A  
; CURRENT FILING DATE: 2003-06-09  
; PRIOR APPLICATION NUMBER: 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: 10/165,250  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: 60/347,062  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-10-457-614A-20

Query Match 14.4%; Score 156; DB 15; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 37 TNEHDLAVHKHKHEMTLKFGRNDSVIV 65  
|||||  
Db 1 TNEHDLAVHKHKHEMTLKFGRNDSVIV 29

RESULT 5  
US-10-924-028-20  
; Sequence 20, Application US/10924028  
; Publication No. US20050043241A1  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501 DIV  
; CURRENT APPLICATION NUMBER: US/10/924,028  
; CURRENT FILING DATE: 2004-08-23  
; PRIOR APPLICATION NUMBER: US/09/970,515  
; PRIOR FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: USSN 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-10-924-028-20

Query Match 14.4%; Score 156; DB 17; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.2e-07;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 37 TNEHDLAVHKHKHEMTLKFGRNDSVIV 65  
|||||  
Db 1 TNEHDLAVHKHKHEMTLKFGRNDSVIV 29

RESULT 6  
US-10-115-178-6  
; Sequence 6, Application US/10115178  
; Publication No. US20020119135A1



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 03:11:51 ; Search time 33 Seconds  
(without alignments)  
472.777 Million cell updates/sec

Title: Aah26175  
Perfect score: 1084  
Sequence: 1 MKFKLVNSARQYKDLNMS.....PSTSTVTITQAPSSNRPIV 209

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1084	100.0	501	4	US-09-949-016-11281
2	257	23.7	52	4	US-09-513-999C-7023
3	156	14.4	29	4	US-09-503-954A-20
4	156	14.4	29	4	US-09-970-515A-20
5	144	13.3	27	3	US-08-819-177-6
6	112	10.3	447	3	US-09-234-332-11
7	106	9.8	443	4	US-09-342-325C-42
8	104.5	9.6	700	4	US-09-976-594-67
9	104.5	9.6	700	4	US-09-538-092-893
10	102.5	9.5	496	2	US-08-224-482-2
11	102.5	9.5	533	1	US-08-040-484-1
12	102.5	9.5	533	1	US-08-466-344-1
13	102.5	9.5	533	4	US-09-546-013-21
14	102.5	9.5	533	6	5206152-2
15	102.5	9.5	533	6	5206152-2
16	101	9.3	878	4	US-09-556-706B-2
17	101	9.3	878	4	US-09-724-418A-2
18	101	9.3	907	3	US-08-783-774-2
19	101	9.3	907	4	US-09-328-599A-1
20	101	9.3	907	5	PCT-US95-04611A-19
21	100.5	9.3	501	4	US-09-342-325C-44
22	100	9.2	97	4	US-09-395-448-15
23	100	9.2	97	4	US-09-925-796-15
24	100	9.2	97	4	US-09-941-450-15
25	100	9.2	1557	4	US-09-410-551B-27
26	100	9.2	1557	4	US-09-940-316B-27
27	97.5	9.0	306	4	US-10-029-180-60

28	96	8.9	522	4	US-09-949-016-11145	Sequence 11145, A
29	96	8.9	667	4	US-09-949-016-7759	Sequence 7759, Ap
30	96	8.9	753	4	US-09-949-016-6676	Sequence 6676, Ap
31	95.5	8.8	556	4	US-09-949-016-7134	Sequence 7134, Ap
32	95.5	8.8	613	4	US-09-438-833-6	Sequence 6, Appli
33	95	8.8	1140	4	US-09-538-092-647	Sequence 647, App
34	94.5	8.7	311	4	US-09-438-833-7	Sequence 7, Appli
35	94.5	8.7	532	4	US-09-949-016-7389	Sequence 7389, Ap
36	94.5	8.7	652	4	US-09-438-833-5	Sequence 5, Appli
37	94.5	8.7	756	4	US-09-438-833-11	Sequence 11, Appl
38	94.5	8.7	805	2	US-08-480-473B-4	Sequence 4, Appli
39	94.5	8.7	805	3	US-08-915-213-4	Sequence 4, Appli
40	94.5	8.7	805	3	US-09-235-217-4	Sequence 4, Appli
41	94.5	8.7	805	5	PCT-US96-10251-4	Sequence 4, Appli
42	94.5	8.7	813	4	US-09-438-833-12	Sequence 12, Appl
43	94.5	8.7	826	1	US-08-785-241-6	Sequence 6, Appli
44	94.5	8.7	826	2	US-08-480-473B-2	Sequence 2, Appli
45	94.5	8.7	826	3	US-08-915-213-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-949-016-11281  
; Sequence 11281, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11281  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-11281

Query Match	100.0%;	Score 1084;	DB 4;	Length 501;
Best Local Similarity	100.0%;	Pred. No. 1.1e-106;		
Matches 209;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKFKLVNSARQYKDLNMSDDKPFCTAGCGQRTNEDHLAVHKHMTLKFGPARN	60	
Db	5	MKFKLVNSARQYKDLNMSDDKPFCTAGCGQRTNEDHLAVHKHMTLKFGPARN	64	
QY	61	DSVIVADQPTPTRFKLNCEEVGLFNELASPFNEFKKASEDDIKMPLDLSPLATPIIR	120	
Db	65	DSVIVADQPTPTRFKLNCEEVGLFNELASPFNEFKKASEDDIKMPLDLSPLATPIIR	124	
QY	121	SKIEEPSVVEVTHQDSPLPHPESTTTSDEKEVPLAQTAQPTSAIVRPASLQVNVLTSSD	180	
Db	125	SKIEEPSVVEVTHQDSPLPHPESTTTSDEKEVPLAQTAQPTSAIVRPASLQVNVLTSSD	184	
QY	181	SSVIIQOAVPSPTSTVTITQAPSSNRPIV	209	
Db	185	SSVIIQOAVPSPTSTVTITQAPSSNRPIV	213	

RESULT 2  
US-09-513-999C-7023  
; Sequence 7023, Application US/09513999C  
; Patent No. 6783961

GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59 US2 REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7023  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 6  
; OTHER INFORMATION: Xaa=Pro or Ser  
US-09-513-999C-7023

Query Match 23.7%; Score 257; DB 4; Length 52;  
Best Local Similarity 86.5%; Pred. No. 3.9e-20;  
Matches 45; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 19 MSDDKPFLLCTAGCGQRTNEDHLAVHKHKHMTLKFPGPARNDSDVIVADQTP 70  
DB 1 MGDDRXFVCNAPGCGQRTNEDHLAVHKHKHMTLKFPGPARNDSDVIVADQTP 52

RESULT 3  
US-09-503-954A-20  
; Sequence 20, Application US/09503954A  
; Patent No. 6610820  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE OF INVENTION: TRANSDUCTION PATHWAY  
; FILE REFERENCE: 20349-501  
; CURRENT APPLICATION NUMBER: US/09/503,954A  
; CURRENT FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: USSN 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-09-503-954A-20

Query Match 14.4%; Score 156; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 9.1e-10;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEDHLAVHKHKHMTLKFPGPARNDSDVIV 65  
DB 1 TNEDHLAVHKHKHMTLKFPGPARNDSDVIV 29

RESULT 4  
US-09-970-515A-20  
; Sequence 20, Application US/09970515A  
; Patent No. 6780970  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE OF INVENTION: TRANSDUCTION PATHWAY  
; FILE REFERENCE: 20349-501 DIV

; CURRENT APPLICATION NUMBER: US/09/970,515A  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-09-970-515A-20

Query Match 14.4%; Score 156; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 9.1e-10;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEDHLAVHKHKHMTLKFPGPARNDSDVIV 65  
DB 1 TNEDHLAVHKHKHMTLKFPGPARNDSDVIV 29

RESULT 5  
US-08-819-177-6  
; Sequence 6, Application US/08819177  
; Patent No. 6043083  
; GENERAL INFORMATION:  
; APPLICANT: Davis, Roger J.  
; APPLICANT: Dickens, Martin  
; TITLE OF INVENTION: INHIBITORS OF THE JNK SIGNAL  
; TITLE OF INVENTION: TRANSDUCTION PATHWAY AND METHODS OF USE  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/819,177  
; FILING DATE: 28 April 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. Peter  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/037001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION: ATF-2 JNK-binding domain  
US-08-819-177-6

Query Match 13.3%; Score 144; DB 3; Length 27;  
Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 03:06:25 ; Search time 96 Seconds  
(without alignments)  
842.010 Million cell updates/sec

Title: AAH26175

Perfect score: 1084

Sequence: 1 MKFKLVNSAQYKDLWNMS.....PSPTSTVITQAPSSNRPIV 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_16Dec04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1980s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	940.5	86.8	482	8 ADR10403	ADR10403 Human pro
2	571	52.7	413	8 ADP80860	ADP80860 Mouse C13
3	566	52.2	411	5 ABB57020	ABB57020 Mouse isc
4	390.5	36.0	501	7 ADC18733	ADC18733 Human cyc
5	257	23.7	52	3 AAG02942	AAG02942 Human sec
6	171	15.8	359	4 ABB67482	ABB67482 Drosophil
7	156	14.4	29	4 AAB668325	AAB668325 c-Jun ami
8	156	14.4	29	7 ADF17925	ADF17925 Synthetic
9	156	14.4	29	8 ADO01349	ADO01349 c-Jun ami
10	112	10.3	447	3 AAB14349	AAB14349 Human Zic
11	112	10.3	447	8 ADN04236	ADN04236 Antipsori
12	112	10.3	447	8 ADQ17360	ADQ17360 Human sof
13	110.5	10.2	1292	8 ADQ39909	ADQ39909 Human myo
14	110	10.1	447	5 ABB57042	ABB57042 Mouse isc
15	109	10.1	752	4 ABB60124	ABB60124 Drosophil
16	107.5	9.9	791	4 ABB67268	ABB67268 Drosophil
17	107.5	9.9	845	4 ABB64280	ABB64280 Drosophil
18	106.5	9.8	1292	7 ABB31295	ABB31295 Human nov
19	106.5	9.8	1484	4 ABG18797	ABG18797 Novel hum
20	105	9.7	2992	8 ADP30190	ADP30190 Human sec
21	105	9.7	3065	8 ADP30259	ADP30259 Human sec
22	105	9.7	3394	7 ADJ68723	ADJ68723 Human hea
23	104.5	9.6	512	8 ADP22648	ADP22648 Sea-squir
24	104.5	9.6	700	4 AAM40233	AAM40233 Human pol
25	104.5	9.6	700	6 ABR92125	ABR92125 Human cer

26	104.5	9.6	700	7 ADD14182	ADD14182 Human src
27	104.5	9.6	700	7 ADN95317	ADN95317 Human BEC
28	104.5	9.6	700	8 ADL12338	ADL12338 Human ste
29	104.5	9.6	700	8 ADL70593	ADL70593 Cervical
30	104.5	9.6	700	8 ABM82242	ABM82242 Tumour-as
31	104.5	9.6	724	4 AAM42019	AAM42019 Human pol
32	104	9.6	1982	4 ABG16404	ABG16404 Novel hum
33	104	9.6	1982	4 ABG19656	ABG19656 Novel hum
34	104	9.6	1983	8 ADQ19071	ADQ19071 Human sof
35	104	9.6	2768	4 ABB68397	ABB68397 Drosophil
36	103.5	9.5	929	8 ADR08894	ADR08894 Human pro
37	103.5	9.5	1483	8 ADQ97206	ADQ97206 Mouse can
38	103	9.5	1044	5 ABP64908	ABP64908 Human pro
39	103	9.5	1166	7 ADB64795	ADB64795 Human pro
40	103	9.5	4873	6 ABO14747	ABO14747 Novel hum
41	102.5	9.5	442	7 ADE63885	ADE63885 Rat Prote
42	102.5	9.5	496	2 AAW76984	AAW76984 Mouse Egr
43	102.5	9.5	533	1 AAP93113	AAP93113 Egr-1, 9/
44	102.5	9.5	533	2 AAR63129	AAR63129 Mouse Egr
45	102.5	9.5	533	4 AAG78054	AAG78054 Human zfin

#### ALIGNMENTS

##### RESULT 1

ADR10403

ID ADR10403 standard; protein; 482 AA.

AC ADR10403;

XX 04-NOV-2004 (first entry)

DE Human protein useful for treating neurological disease Seq 3909.

XX human; oligo-capping method; diagnostic marker; gene therapy;

KW osteoporosis; neurological disease; Alzheimer's disease;

KW Parkinson's disease; dementia; short memory; cancer;

KW sense or motor function; emotional reaction; fear response; panic;

KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;

KW tranquiliser.

OS Homo sapiens.

XX EPI447413-A2.

PN 18-AUG-2004.

PD 12-FEB-2004; 2004EP-00003145.

PF 14-FEB-2003; 2003JP-00102207.

PR 09-MAY-2003; 2003JP-00131452.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;

PI Wakamatsu A, Ishii S, Nagai K, Irie R;

DR WPI; 2004-583265/57.

XX N-PSDB; ADR08447.

XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases, Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

PS Claim 1; SEQ ID NO 3909; 2686pp; English.

XX This invention relates to novel, isolated full length human cDNA

XX molecules and the encoded proteins thereof. Specifically, it refers to

XX cDNA clones obtained by an oligo-capping method, where none of these

XX clones are identical to any known human mRNAs. The present invention

XX describes an immunoassay to identify agonists and antagonists, as well as

XX antibodies, antisense molecules and siRNAs that can all be used to bind

XX to and modulate expression of the cDNA molecules. As such, these

CC molecules are useful for diagnostic markers or therapeutic targets for  
 CC the various diseases or morbid states. In particular, they are useful in  
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
 CC disease, Parkinson's disease, dementia, short memory and various cancers,  
 CC as well as for maintaining equilibrium of sense or motor function, and  
 CC for treating emotional reaction, fear response and panic. Accordingly,  
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
 CC cytoskeletal and tranquiliser activities. This polypeptide is a protein  
 CC encoded by a full length human cDNA sequence of the invention. NOTE: This  
 CC sequence is not given in the sequence listing of the specification but  
 CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-  
 CC office.

XX SQ Sequence 482 AA;

Query Match 86.8%; Score 940.5; DB 8; Length 482;  
 Best Local Similarity 89.0%; Pred. No. 2.8e-79;  
 Matches 186; Conservative 0; Mismatches 0; Indels 23; Gaps 1;

QY 1 MKFKLHVSARQYKOLWNMSDDKPFCTAPGCGQRTNEDHLAVHKHMTLKEFGPARN 60  
 DB 1 MKFKLHVSARQYKOLWNMSDDKPFCTAPGCGQRTNEDHL----- 42  
 QY 61 DSVIVADOTPTTRFLKNCVEEVLNGLASPFENEFKASDDIKKMPDLPLATPIIR 120  
 DB 43 -----ADQPTPTTRFLKNCVEEVLNGLASPFENEFKASDDIKKMPDLPLATPIIR 97  
 QY 121 SKIEPVSVVETTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAVRPASLQVNPVLLTSSD 180  
 DB 98 SKIEPVSVVETTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAVRPASLQVNPVLLTSSD 157  
 QY 181 SSVIIQAVPSPSTSTVITQAPSSNRPIV 209  
 DB 158 SSVIIQAVPSPSTSTVITQAPSSNRPIV 186

RESULT 2

ADP80860  
 ID ADP80860 standard; protein; 413 AA.

XX AC ADP80860;

XX 09-SEP-2004 (first entry)

XX DE Mouse Cl30020M04Rik amino acid sequence SEQ ID NO:85.

XX KW c-fos; c-fos interacting protein; fos interacting protein chromosome X;  
 KW Fip-cx; screening; Fc-fos interacting protein inhibitor; mouse.

XX OS Mus musculus.

XX PN WO2004053121-A1.

XX PD 24-JUN-2004.

XX PF 19-NOV-2003; 2003WO-JP014749.

XX PR 11-DEC-2002; 2002JP-00360046.

XX PA (UYKE-) UNIV KEIO.

XX PI Miyamoto E, Ishizaka M, Yanagawa H;

XX DR WPI; 2004-517250/49.

XX DR N-PSDB; ADP80924.

XX PT New proteins that interact with fos, e.g., fos interacting protein  
 PT chromosome X (Fip-cx).

XX PS Claim 57; SEQ ID NO 85; 192pp; Japanese.

XX CC The present invention describes a protein (I) that interacts with c-fos  
 CC (e.g., fos interacting protein chromosome X (Fip-cx), Fip-cx.1, fos

CC interacting protein chromosome ex.2, or fos interacting protein  
 CC chromosome 4). Also described: (I) a nucleic acid (II) that encodes (I);  
 CC (2) an inhibitor (III) that inhibits the interaction of (I) encoded by  
 CC (II) with c-fos protein; (3) an inhibitor (IV) that inhibits interaction  
 CC of a protein with c-fos protein; and (4) detecting (MI) the interaction  
 CC of a protein as mentioned in (3) with c-fos protein, by contacting the  
 CC protein with c-fos, to form a composite. (I) is useful for detecting its  
 CC interaction with c-fos, to form a complex. (MI) is useful for screening  
 CC the protein that interacts with c-fos which involves performing the  
 CC detection process and selecting the protein that interacts with c-fos.  
 CC (MI) is useful for screening a protein that interacts with c-fos which  
 CC involves performing (MI) and selecting the detected protein. (I) is  
 CC useful for screening inhibitors that interact with c-fos. The present  
 CC sequence represents a mouse Cl30020M04Rik amino acid sequence, which can  
 CC interact with c-fos in the exemplification of the present invention.

XX SQ Sequence 413 AA;

Query Match 52.7%; Score 571; DB 8; Length 413;  
 Best Local Similarity 63.8%; Pred. No. 9.6e-45;  
 Matches 125; Conservative 16; Mismatches 39; Indels 16; Gaps 6;

QY 19 MSDDKPFCTAPGCGQRTNEDHLAVHKHMTLKEFGPARNDSVIVADQPTPTTRFLKN 78  
 DB 1 MGDPRPFVCSAPGCGQRTNEDHLAVHKHMTLKEFGPARTDSVIVADQPTPTTRFLKN 60  
 QY 79 CEEVGLFNELASPFENEFKASDDIKK---MPLDLSPLATPIIRSKIEPVSVVETTHOD 135  
 DB 61 CEEVGLFNELASSPEHEPFKASDDDEKKAAGPLDMSLPSTPDIKKEEPVEVDSSPPD 120  
 QY 136 SPLPHPESTTSDEKEV---PLAQTAQTSIAVRPASLQVNPVLLTSSDSSVIIQAVPSP 192  
 DB 121 SPASSPCSPPLKEKEVTKPVV-ISTPTPTTVRPGSLP-----LHLGYDP---LHPTLPSP 172  
 QY 193 TSSVTITQAPSSNRPI 208  
 DB 173 TS--VITQAPSPNRQI 186

RESULT 3

ABB57020

ID ABB57020 standard; protein; 411 AA.

XX AC ABB57020;

XX 07-MAR-2002 (first entry)

XX DE Mouse ischaemic condition related protein sequence SEQ ID NO:2.

XX KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

XX OS Mus musculus.

XX PN WO2001188188-A2.

XX PD 22-NOV-2001.

XX PF 18-MAY-2001; 2001WO-JP004192.

XX PR 18-MAY-2000; 2000JP-00145977.

XX PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX DR WPI; 2002-034733/04.

XX DR N-PSDB; ASI99202.

XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 11:19:57 ; Search time 91 Seconds  
(without alignments)  
1176.094 Million cell updates/sec

Title: AAH26175  
Perfect score: 209  
Sequence: 1 MKFKLVNSAQYKDLNMS.....PSPTSSTVITQAPSSNRPIV 209

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot 03:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	209	Q8TAR1	Q8TAR1 homo sapien
2	191	91.4	234	Q8BN75	Q8BN75 mus musculus
3	191	91.4	420	Q8CBR9	Q8CBR9 mus musculus
4	191	91.4	487	1 ATF2 HUMAN	P15336 homo sapien
5	191	91.4	487	1 ATF2 MOUSE	P16951 mus musculus
6	191	91.4	487	2 Q8CBG4	Q8CBG4 mus musculus
7	143	68.4	440	2 Q68FE3	Q68FE3 mus musculus
8	127	60.8	487	1 ATF2 RAT	Q00969 rattus norv
9	102	48.8	487	1 ATF2 CHICK	Q93602 gallus gall
10	62	29.7	486	2 Q91576	Q91576 xenopus lae
11	34	16.3	144	2 Q96J78	Q96J78 homo sapien
12	31	14.8	307	2 Q9UD83	Q9UD83 homo sapien
13	31	14.8	413	2 Q8R0S1	Q8R0S1 mus musculus
14	31	14.8	494	1 ATF7 HUMAN	P17544 homo sapien
15	29	13.9	117	2 Q8IVR8	Q8IVR8 homo sapien
16	19	9.1	135	2 Q8BKN9	Q8BKN9 mus musculus
17	19	9.1	144	2 Q8BM42	Q8BM42 m mus muscu
18	19	9.1	148	2 Q75N02	Q75N02 homo sapien
19	19	9.1	508	1 C8B5 HUMAN	Q2930 homo sapien
20	9	4.3	850	2 Q9W0Z5	Q9W0Z5 drosophila
21	8	3.8	141	2 Q8YWL5	Q8YWL5 anabaena sp
22	8	3.8	312	2 Q6Z519	Q6Z519 caenorhabdi
23	8	3.8	339	2 Q6DC87	Q6DC87 xenopus lae
24	8	3.8	340	2 Q7V1Y2	Q7V1Y2 prochloroco
25	8	3.8	384	2 Q73DB9	Q73DB9 bacillus ce
26	8	3.8	393	2 Q7P753	Q7P753 fusobacteri
27	8	3.8	395	2 Q978Z6	Q978Z6 thermoplas
28	8	3.8	458	2 Q8I5J7	Q8I5J7 plasmodium
29	8	3.8	506	2 Q6ZJEL	Q6ZJEL oryza sativ
30	8	3.8	575	2 Q6LGB3	Q6LGB3 photobacter
31	8	3.8	594	2 Q73932	Q73932 xenopus lae

Query Match

100.0% ; Score 209 ; DB 2 ; Length 209 ;

32 8 3.8 599 2 Q93CJ2 Q93CJ2 anabaena va  
33 8 3.8 887 2 Q9MAQ3 Q9MAQ3 arabidopsis  
34 8 3.8 887 2 Q9ZSB6 Q9ZSB6 arabidopsis  
35 8 3.8 910 2 Q93Y01 Q93Y01 arabidopsis  
36 8 3.8 928 2 Q9T0B6 Q9T0B6 arabidopsis  
37 8 3.8 937 2 Q9T0B8 Q9T0B8 arabidopsis  
38 8 3.8 951 2 Q8H557 Q8H557 oryza sativ  
39 8 3.8 1028 2 Q9ZSB5 Q9ZSB5 arabidopsis  
40 8 3.8 1212 2 Q7RJZ3 Q7RJZ3 plasmodium  
41 8 3.8 1350 2 Q7T3T6 Q7T3T6 brachydanio  
42 8 3.8 1476 2 Q9GRN0 Q9GRN0 leishmania  
43 7 3.3 55 2 Q653G0 Q653G0 oryza sativ  
44 7 3.3 105 2 Q8C8W1 Q8C8W1 mus musculus  
45 7 3.3 116 2 Q6Z8K8 Q6Z8K8 oryza sativ

## ALIGNMENTS

RESULT 1  
Q8TAR1 PRELIMINARY; PRT; 209 AA.  
AC Q8TAR1; 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE ATF2 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore J., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullan S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Villalón D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalka U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RA "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
EMBL; BC026175; AAH26175.1; -  
HSSP; P15336; 1BHI.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 1.  
DR SMART; SM00355; Znf\_C2H2; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 1.  
SQ SEQUENCE 209 AA; 23050 MW; A26AF07CA5D8D5E7 CRC64;

```
Best Local Similarity 100.0%; Pred. No. 1.8e-208;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFKLHVSARQYKDLWNMSDDKFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARN 60
DB 1 MKFKLHVSARQYKDLWNMSDDKFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARN 60

QY 61 DSVIVADQPTPTTRFLKNCCEVGLFNELASPPENEFKASEDDIKMPLDLSPLATPIIR 120
DB 61 DSVIVADQPTPTTRFLKNCCEVGLFNELASPPENEFKASEDDIKMPLDLSPLATPIIR 120

QY 121 SKIEPSSVVEVTHQSDPLPHPESTTSDEKEVPLAQTPSAIVRPASIQVNPVLLTSSD 180
DB 121 SKIEPSSVVEVTHQSDPLPHPESTTSDEKEVPLAQTPSAIVRPASIQVNPVLLTSSD 180

QY 181 SSVIIQQAVPSPSTSTVTITQAPSSNRPIV 209
DB 181 SSVIIQQAVPSPSTSTVTITQAPSSNRPIV 209

RESULT 2
Q8BN75 PRELIMINARY; PRT; 234 AA.
AC Q8BN75;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
DE library, clone:EI3020102 product:activating transcription factor 2,
DE full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN [3]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN [4]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugihara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RN [5]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Yamaoka S., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RN [6]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Kato H., Kawai J., Kohji Y., Itoh M., Kagawa T., Kasukawa T.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numata R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akai H., Tanaka T., Tanaka T.,
RA Tomaru A., Toyama T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK087424; BAC39869.1; -.
DR HSSP; P15336; 1BHI.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
SQ SEQUENCE 234 AA; 25208 MW; F06750FA9EB37A4D CRC64;
```

```
Query Match 91.4%; Score 191; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 1e-189;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 MSDDKPFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARNDSVIVADQPTPTTRFLKN 78
DB 1 MSDDKPFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARNDSVIVADQPTPTTRFLKN 60

QY 79 CEEVGLFNELASPPENEFKASEDDIKMPLDLSPLATPIIRSKIEPSSVVEVTHQSDPL 138
DB 61 CEEVGLFNELASPPENEFKASEDDIKMPLDLSPLATPIIRSKIEPSSVVEVTHQSDPL 120

QY 139 PHPESTTSDEKEVPLAQTPSAIVRPASIQVNPVLLTSSDSSVIIQQAVPSPSTSTVI 198
DB 121 PHPESTTSDEKEVPLAQTPSAIVRPASIQVNPVLLTSSDSSVIIQQAVPSPSTSTVI 180

QY 199 TQAPSSNRPIV 209
DB 181 TQAPSSNRPIV 191

RESULT 3
Q8CBR9 PRELIMINARY; PRT; 420 AA.
AC Q8CBR9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male urinary bladder cDNA, RIKEN full-length
DE enriched library, clone:9530046122 product:activating transcription
DE factor 2, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 14:06:13 ; Search time 3168 seconds  
(without alignments)  
2511.184 Million cell updates/sec

Title: AAH26175

Perfect score: 209

Sequence: 1 MKFKLVNSARQYKDLWNMS.....PSPTSTVTITQAPSSNRPIV 209

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68477535

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q/cgn2 1/USPTO.spool/AAH26175/runat 07072005 101848 20015/app.query.fasta\_1.391  
-DB=EST -OPMT=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 @CGN 1 1 4352 @runat 07072005 101848 20015 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199	95.2	904	7	CR763073
2	195	93.3	1606	3	AK087424 Mus muscu
3	195	93.3	3623	3	AK035438 Mus muscu
4	193	92.3	900	5	B0158697
5	193	92.3	956	5	B0158697
6	178	85.2	883	4	B1854668
7	170	81.3	650	7	CR767893
8	166	79.4	792	6	CA750476
9	164	78.5	613	5	BU662818 cl91g01.2

10	164	78.5	802	7	CR629564
11	163	78.0	670	7	CR367891
12	153	73.2	746	6	CD348287
13	153	73.2	749	6	CD348288
14	151	72.2	915	5	B0715489
15	149	71.3	806	6	CB961368
16	147	70.3	664	7	CF535749
17	145	69.4	577	7	CR367889
18	145	69.4	651	7	CR367890
19	143	68.4	686	2	BE882603
20	143	68.4	757	7	CK635513
21	143	68.4	875	5	B0891414
22	143	68.4	2106	3	AK051591
23	143	68.4	3579	3	AK041145
24	142	67.9	552	9	CL570065
25	138	66.0	780	7	CF742229
26	135	64.6	763	7	CR367888
27	135	64.6	958	5	BU503620
28	134	64.1	417	8	CC249614
29	133	63.6	818	7	CK455466
30	132	63.2	742	7	CK634629
31	132	63.2	758	6	CA510974
32	132	63.2	781	6	CA512317
33	132	63.2	2835	3	AY724488
34	122	58.4	621	3	CR367892
35	115	55.5	585	1	AI159526
36	115	55.0	700	5	BU611698
37	109	52.2	924	4	B1551997
38	108	51.7	723	7	CK454827
39	107	51.2	722	7	CF532094
40	106	50.7	400	2	BE818106
41	106	50.7	606	5	BQ037883
42	106	50.7	640	7	CK455108
43	106	50.7	744	1	AJ453735
44	106	50.7	826	1	AJ454852
45	104	49.8	576	5	BP257160

ALIGNMENTS

RESULT 1

CR763073

LOCUS

DEFINITION

CR763073 904 bp mRNA linear EST 23-SEP-2004  
DKFZp46911824 r1 469 (synonym: pkid1) Pongo pygmaeus cdna clone  
DKFZp46911824-5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR763073.1 GI:52600802

EST.

Pongo pygmaeus (orangutan)

Pongo pygmaeus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.

1 (bases 1 to 904)

Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,

Mewes,H.W., Weill,B., Amid,C., Osanger,A., Fobo,G., Han,M. and

Wiemann,S.

Pongo pygmaeus mRNA (Ansorge,W., Krieger,S., Regiert,T., et al.)

Unpublished (2004)

Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert. Clone from S. Wiemann,

Molecular Genome Analysis, German Cancer Research Center (DKFZ);

Email s.wiemann@dkfz-heidelberg.de; rlin, Germany. Please contact

RZPD for ordering:

http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp46911824

Further information about the clone and the sequencing project is

available at http://mips.gsf.de/projects/cdna/.

Location/Qualifiers

1..904

/organism="Pongo pygmaeus"

/mol\_type="mRNA"

/db\_xref="taxon:9609"

```
/clone="DKFZp46911824"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="469 (synonym: pkidi)"
/note="Vector: pSport1_Sfi; Site_1: SfiI; Site_2: SfiIb"

ORIGIN

Alignment Scores:
Pred. No.: 4,078-186 Length: 904
Score: 199.00 Matches: 199
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 95.22% Indels: 0
DB: 7 Gaps: 0

AAH26175 (1-209) x CR763073 (1-904)

QY 1 MetLysPheLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20
DB 273 ATGAATTCACAGTTACATGTGAATTCGCCAGGCATACAGGACCTGTGGAATATGAGT 332
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40
DB 333 GATGACAAACCCCTTCTATGCACTCGGCTGGATGTGGCCAGCGTTTACCAAGGAGAT 392
QY 41 HisLeuAlaValHisGlyHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60
DB 393 CATTTGGCTGTCCATAAAACATAAATGAGATGACACTGAAATTTGGTCCAGCAGCTAAT 452
QY 61 AspSerValIleValAlaAspGlnThrProThrProThrArgPheLeuLysAsnCysGlu 80
DB 453 GACAGTGCATTTGGTGGTGGTATCAGACCCCAACCAAGATTTCTTGAAAACCTGTGAA 512
QY 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100
DB 513 GAAGTGGGTGTGTTAATGAGTTGGGAGTTCATTTGAGATGAATTCAGAAAGCTTCA 572
QY 101 GluAspAspIleLysGlyMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120
DB 573 GAAGACGACATTAAAAAATGCTCTAGATTTATCCCTCTTGCACACCTATCATAGA 632
QY 121 SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140
DB 633 AGCAAAATTTAGGAGCCCTTCTGTTAGAAAGCATCCACGAGTAGTCTTTTACCTCAC 692
QY 141 ProGluSerThrThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr 160
DB 693 CCAGAGTCTACTACAGTGATGAGAGGAGTAGTACCACTGGCACAACCTGCACAGCCACG 752
QY 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180
DB 753 TCAGCTATTGTTGCTCCAGCATCATACAGGTTCCCAATGTGCTGCTTACAAGTTCTGAC 812
QY 181 SerSerValIleIleGlnGlnAlaValProSerProThrSerSerThrValIleThr 199
DB 813 TCAAGTGAATTAATTATTCAGAGGAGTACCTTCCACCAACCTCAAGTACTGTAATCACC 869

RESULT 2
AK087424
LOCUS
DEFINITION Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
2, full insert sequence.
ACCESSION AK087424
VERSION AK087424.1 GI:26352477
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
```

High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
9279253  
10349636

2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 1606)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.  
Location/Qualifiers  
1. 1606  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:E130020102"  
/db\_xref="taxon:10090"  
/clone="E130020102"

FEATURES  
source





QY 121 SKIEPSVVTTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAIVRPASLOVNVLLTSSD 180  
DB 121 SKIEPSVVTTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAIVRPASLOVNVLLTSSD 180  
QY 181 SSVIIQQAVPSPTSSVTITQAPSSNRPIV 209  
DB 181 SSVIIQQAVPSPTSSVTITQAPSSNRPIV 209

RESULT 2  
B42026  
cyclic AMP response element DNA-binding protein isoform 1 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Aug-1999  
C:Accession: B42026  
R:Georgopoulos, K.; Morgan, B.A.; Moore, D.D.  
Mol. Cell. Biol. 12, 747-757, 1992  
A:Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activation of transcription  
A:Reference number: A42026; MUID:92123199; PMID:1531087  
A:Contents: EL4  
A:Accession: B42026  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-456 <GEO>  
A:Cross-references: GB:S76657; NID:9243428; PIDN:AAB21128.1; PID:9243429  
A:Note: sequence extracted from NCBI backbone (NCBIN:76657, NCBIP:76658)  
C:Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homology F:297-337/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 76.1%; Score 159; DB 2; Length 456;  
Best Local Similarity 100.0%; Pred. No. 1.1e-160; Mismatches 0; Indels 0; Gaps 0;  
Matches 159; Conservative 0;

QY 51 MTLKFGPARNDSVIVADQPTTFLKNCCEEVGLFNELASPFENFKKASEDDIKKMPLD 110  
DB 1 MTLKFGPARNDSVIVADQPTTFLKNCCEEVGLFNELASPFENFKKASEDDIKKMPLD 60

QY 111 LSPLATPIIRSKIEPSVVTTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAIVRPASLIQ 170  
DB 61 LSPLATPIIRSKIEPSVVTTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAIVRPASLIQ 120

QY 171 VNVLLTSSDSSVIIQQAVPSPTSSVTITQAPSSNRPIV 209  
DB 121 VNVLLTSSDSSVIIQQAVPSPTSSVTITQAPSSNRPIV 159

RESULT 3  
A42026  
cyclic AMP response element-binding protein 3 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 26-Aug-1999  
C:Accession: A42026  
R:Georgopoulos, K.; Morgan, B.A.; Moore, D.D.  
Mol. Cell. Biol. 12, 747-757, 1992  
A:Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activation of transcription  
A:Reference number: A42026; MUID:92123199; PMID:1531087  
A:Contents: EL4  
A:Accession: A42026  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-448 <GEO>  
A:Cross-references: GB:S76655; NID:9243426; PIDN:AAB21127.1; PID:9243427  
A:Note: sequence extracted from NCBI backbone (NCBIN:76655, NCBIP:76656)  
C:Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homology F:289-329/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 68.4%; Score 143; DB 2; Length 448;  
Best Local Similarity 100.0%; Pred. No. 1.1e-143; Mismatches 0; Indels 0; Gaps 0;  
Matches 143; Conservative 0;

QY 67 DOTPTTFLKNCCEEVGLFNELASPFENFKKASEDDIKKMPLDLSPLATPIIRSKIEP 126  
|||||

DB 9 DOTPTTFLKNCCEEVGLFNELASPFENFKKASEDDIKKMPLDLSPLATPIIRSKIEP 68  
QY 127 SVVETTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAIVRPASLOVNVLLTSSDSSVIIQ 186  
DB 69 SVVETTHQDSPLPHPESTTSDEKEVPLAQTAQTSIAIVRPASLOVNVLLTSSDSSVIIQ 128  
QY 187 QAVPSPTSSVTITQAPSSNRPIV 209  
DB 129 QAVPSPTSSVTITQAPSSNRPIV 151

RESULT 4  
A39429  
cAMP response element-binding protein ATF2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: A39429  
R:Kageyama, R.; Sasai, Y.; Nakanishi, S.  
J. Biol. Chem. 266, 15525-15531, 1991  
A:Title: Molecular characterization of transcription factors that bind to the cAMP response element  
A:Reference number: A39429; MUID:91332085; PMID:1714459  
A:Accession: A39429  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-389 <KAG>  
A:Cross-references: UNIPROT:Q00969; GB:M65148; NID:g206569; PIDN:AAA42013.1; PID:g206570  
A:Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homology F:231-271/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 60.8%; Score 127; DB 1; Length 389;  
Best Local Similarity 100.0%; Pred. No. 1e-126; Mismatches 0; Indels 0; Gaps 0;  
Matches 127; Conservative 0;

QY 19 MSDDKPFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARNDSVIVADQPTTFLKLN 78  
DB 1 MSDDKPFLCTAPGCGQRTNEDHLAVHKHHEMTLKFGPARNDSVIVADQPTTFLKLN 60

QY 79 CEEVGLFNELASPFENFKKASEDDIKKMPLDLSPLATPIIRSKIEPSVVTTHQDSPL 138  
DB 61 CEEVGLFNELASPFENFKKASEDDIKKMPLDLSPLATPIIRSKIEPSVVTTHQDSPL 120

QY 139 PHPESTT 145  
DB 121 PHPESTT 127

RESULT 5  
C42026  
cyclic AMP response element DNA-binding protein isoform 2 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Aug-1999  
C:Accession: C42026  
R:Georgopoulos, K.; Morgan, B.A.; Moore, D.D.  
Mol. Cell. Biol. 12, 747-757, 1992  
A:Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate activation of transcription  
A:Reference number: A42026; MUID:92123199; PMID:1531087  
A:Contents: EL4  
A:Accession: C42026  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-358 <GEO>  
A:Cross-references: GB:S76659; NID:g243430; PIDN:AAB21129.1; PID:g243431  
A:Note: sequence extracted from NCBI backbone (NCBIN:76659, NCBIP:76660)  
C:Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homology F:199-239/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 47.4%; Score 99; DB 2; Length 358;  
Best Local Similarity 100.0%; Pred. No. 5.5e-97; Mismatches 0; Indels 0; Gaps 0;  
Matches 99; Conservative 0;

QY 51 MTLKFGPARNDSVIVADQPTTFLKNCCEEVGLFNELASPFENFKKASEDDIKKMPLD 110  
|||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 15:27:50 ; Search time 594 Seconds  
(without alignments)  
2209.037 Million cell updates/sec

Title: AAH26175

Perfect score: 209

Sequence: 1 MKFLHVSARQYKDLNMS.....PSPTSTVITQAPSSNRPIV 209

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6330943 seqs, 3139157217 residues

Word size: 1

Total number of hits satisfying chosen parameters: 12650797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/AAH26175/runat\_07072005\_101851\_20099/app\_query.fasta\_1.391  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=oli.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=AAH26175 @cgn 1.1.741 @runat\_07072005\_101851\_20099  
-NCPU=6 -ICPU=3 -NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/FCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10J\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	27.3	501	9	US-09-833-790-215 Sequence 215, App
2	34	16.3	509	16	US-10-210-120-82 Sequence 82, Appl
3	34	16.3	509	22	US-10-909-035-82 Sequence 82, Appl
4	31	14.8	800	20	US-10-357-930-23141 Sequence 23141, A
5	31	14.8	800	20	US-10-357-930-23011 Sequence 29011, A
6	31	14.8	2755	20	US-10-357-930-22633 Sequence 22633, A
7	31	14.8	2755	20	US-10-357-930-28476 Sequence 28476, A
8	19	9.1	2345	11	US-10-723-860-7409 Sequence 7409, Ap
9	19	9.1	2637	11	US-09-968-007A-974 Sequence 974, App
10	19	9.1	2637	15	US-10-171-581-285 Sequence 285, App
11	19	9.1	2637	21	US-10-843-641A-7444 Sequence 7444, Ap
12	17	8.1	51	15	US-10-211-088-92 Sequence 92, Appl
13	16	7.7	51	15	US-10-211-088-80 Sequence 80, Appl
14	14	6.7	273	20	US-10-357-930-4968 Sequence 4968, Ap
15	13	6.2	397	20	US-10-357-930-14137 Sequence 14137, A
16	13	6.2	432	20	US-10-357-930-35258 Sequence 35258, A
17	13	6.2	432	20	US-10-357-930-44098 Sequence 44098, A
C 18	9	4.3	1121	19	US-10-767-701-10951 Sequence 10951, A
C 19	9	4.3	1264	18	US-10-424-599-98072 Sequence 98072, A
C 20	9	4.3	2775	9	US-09-738-626-2994 Sequence 2994, Ap
C 21	9	4.3	2809	21	US-10-494-672-95 Sequence 95, Appl
C 22	9	4.3	254366	10	US-09-822-871-3 Sequence 3, Appl
C 23	9	4.3	254366	18	US-10-673-885-3 Sequence 3, Appl
C 24	9	4.3	3309400	9	US-09-738-626-1 Sequence 1, Appl
C 25	8	3.8	139	9	US-09-728-444-136 Sequence 136, App
C 26	8	3.8	204	18	US-10-424-599-53079 Sequence 53079, A
C 27	8	3.8	275	19	US-10-674-124A-2643 Sequence 2643, Ap
C 28	8	3.8	290	9	US-09-294-093B-2862 Sequence 2862, Ap
C 29	8	3.8	296	9	US-09-867-701-8314 Sequence 8314, Ap
C 30	8	3.8	333	14	US-10-043-875-294 Sequence 294, App
C 31	8	3.8	333	14	US-10-043-875-307 Sequence 307, App
C 32	8	3.8	339	19	US-10-437-963-29382 Sequence 29382, A
C 33	8	3.8	378	20	US-10-425-115-47346 Sequence 47346, A
C 34	8	3.8	396	20	US-10-425-115-3184 Sequence 3184, Ap
C 35	8	3.8	406	19	US-10-437-963-87906 Sequence 87906, A
C 36	8	3.8	419	10	US-09-918-995-9121 Sequence 9121, Ap
C 37	8	3.8	431	9	US-09-764-847-1885 Sequence 1885, Ap
C 38	8	3.8	431	9	US-09-764-847-1886 Sequence 1886, Ap
C 39	8	3.8	431	14	US-10-092-154-1885 Sequence 1885, Ap
C 40	8	3.8	431	14	US-10-092-154-1886 Sequence 1886, Ap
C 41	8	3.8	440	9	US-09-764-847-449 Sequence 449, App
C 42	8	3.8	440	14	US-10-092-154-449 Sequence 449, App
C 43	8	3.8	441	20	US-10-425-115-161483 Sequence 161483, A
C 44	8	3.8	463	9	US-09-954-531-1012 Sequence 1012, Ap
C 45	8	3.8	463	21	US-10-843-641A-2079 Sequence 2079, Ap

# ALIGNMENTS

RESULT 1  
US-09-833-790-215  
; Sequence 215, Application US/09833790  
; Patent No. US20020068288A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Secrist, Heather  
; APPLICANT: Mohamath, Raedoh  
; APPLICANT: Indirias, Carol Y.  
; APPLICANT: Fan, Liqun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.512  
; CURRENT APPLICATION NUMBER: US/09/833,790  
; CURRENT FILING DATE: 2001-04-11

```
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 215
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(501)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-790-215

Alignment Scores:
Pred. No.: 1,15e-48 Length: 501
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.27% Indels: 0
DB: Gaps: 0

AAH26175 (1-209) x US-09-833-790-215 (1-501)

QY 1 MetLysPheLysLeuHisValIasnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20
Db 267 ATGAAATTCAGTTACATGTGAATTCGCGAGCAATACAGGACCTGTGGAATATGAGT 326
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40
Db 327 GATGACAAACCCCTTTCTATGTACTGCGCCTGGATGTGGCAGCGTGTACCAACGAGGAT 386
QY 41 HisLeuAlaValHisIshyHisIshyHisGluMetThrLeuLysPheGlyPro 57
Db 387 CATTTCGCTGTCATAAACAATGAATGATGACACTGAAATTTGGTCCA 437

RESULT 2
US-10-210-120-82
; Sequence 82, Application US/10210120
; Publication No. US20030175736A1
; GENERAL INFORMATION:
; APPLICANT: Chinmalyan, Arul M.
; APPLICANT: Rubin, Mark A.
; APPLICANT: Sreekumar, Arun
; TITLE OF INVENTION: Expression Profile of Prostate Cancer
; FILE REFERENCE: UM-07221
; CURRENT APPLICATION NUMBER: US/10/210,120
; PRIOR FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US 60/309,581
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: US 60/334,468
; PRIOR FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 82
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-210-120-82

Alignment Scores:
Pred. No.: 4,68e-25 Length: 509
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.27% Indels: 0
DB: Gaps: 0

AAH26175 (1-209) x US-10-210-120-82 (1-509)

QY 1 MetLysPheLysLeuHisValIasnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20
Db 1 ATGAAATTCAGTTACATGTGAATTCGCGAGCAATACAGGACCTGTGGAATATGAGT 60
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGln 34
Db 61 GATGACAAACCCCTTTCTATGTACTGCGCCTGGATGTGGCCAG 102

RESULT 3
US-10-909-035-82
; Sequence 82, Application US/10909035
; Publication No. US20050136493A1
; GENERAL INFORMATION:
; APPLICANT: Rubin, Mark A.
; APPLICANT: Chinmalyan, Arul M.
; APPLICANT: Laxman, Bharathi
; APPLICANT: Sreekumar, Arun
; TITLE OF INVENTION: AMACR Cancer Markers
; FILE REFERENCE: UM-09098
; CURRENT APPLICATION NUMBER: US/10/909,035
; CURRENT FILING DATE: 2004-07-30
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 82
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-909-035-82

Alignment Scores:
Pred. No.: 4,68e-25 Length: 509
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.27% Indels: 0
DB: Gaps: 0

AAH26175 (1-209) x US-10-909-035-82 (1-509)

QY 1 MetLysPheLysLeuHisValIasnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20
Db 1 ATGAAATTCAGTTACATGTGAATTCGCGAGCAATACAGGACCTGTGGAATATGAGT 60
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGln 34
Db 61 GATGACAAACCCCTTTCTATGTACTGCGCCTGGATGTGGCCAG 102

RESULT 4
US-10-357-930-23141
; Sequence 23141, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23141
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: July 8, 2005, 14:07:03 ; Search time 183 Seconds  
(without alignments)  
1868.753 Million cell updates/sec

Title: AAh26175  
Perfect score: 209  
Sequence: 1 MFKFLVNSARQYKDLWNMS.....PSTSTVTITQAPSSNRPIV 209

Scoring table:  
OLIGO Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 2400006

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/AAH26175/runat 07072005 101849 20027/app query.fasta\_1.391  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=ol1.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 @CGN 1 1 177 @runat 07072005 101849 20027 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	1621	4	US-09-949-016-5410
2	60	28.7	65966	4	US-09-949-016-17152
3	31	14.8	229	4	US-09-513-999C-2946
C 4	31	14.8	601	4	US-09-949-016-189073
C 5	24	11.5	601	4	US-09-949-016-189064
C 6	11	5.3	601	4	US-09-949-016-189057
7	9	4.3	3346	4	US-09-902-540-548
C 8	9	4.3	157822	4	US-09-949-016-16723
C 9	9	4.3	254366	4	US-09-822-871-3
C 10	8	3.8	158	4	US-09-513-999C-26841
C 11	8	3.8	462	4	US-09-248-796A-8283
12	8	3.8	601	4	US-09-949-016-21288

13	8	3.8	601	4	US-09-949-016-56382	Sequence 56382, A
14	8	3.8	601	4	US-09-949-016-56383	Sequence 56383, A
C 15	8	3.8	601	4	US-09-949-016-83959	Sequence 83959, A
C 16	8	3.8	601	4	US-09-949-016-83960	Sequence 83960, A
17	8	3.8	601	4	US-09-949-016-137071	Sequence 137071, A
18	8	3.8	601	4	US-09-949-016-137072	Sequence 137072, A
19	8	3.8	601	4	US-09-949-016-137073	Sequence 137073, A
C 20	8	3.8	601	4	US-09-949-016-143248	Sequence 143248, A
C 21	8	3.8	601	4	US-09-949-016-143419	Sequence 143419, A
22	8	3.8	601	4	US-09-949-016-163200	Sequence 163200, A
23	8	3.8	601	4	US-09-949-016-179127	Sequence 179127, A
24	8	3.8	601	4	US-09-949-016-191387	Sequence 191387, A
C 25	8	3.8	1001	3	US-09-641-638-205	Sequence 205, App
C 26	8	3.8	1001	3	US-09-641-638-206	Sequence 206, App
C 27	8	3.8	1001	3	US-10-170-097-205	Sequence 205, App
C 28	8	3.8	1001	4	US-10-170-097-206	Sequence 206, App
C 29	8	3.8	1242	4	US-09-489-039A-1286	Sequence 1286, Ap
C 30	8	3.8	1365	4	US-09-857-556A-17	Sequence 17, Appl
C 31	8	3.8	1488	4	US-09-949-016-3851	Sequence 3851, Ap
C 32	8	3.8	2689	2	US-08-876-546A-15	Sequence 15, Appl
C 33	8	3.8	2689	3	US-09-412-252-15	Sequence 15, Appl
C 34	8	3.8	3162	4	US-09-328-352-932	Sequence 932, App
35	8	3.8	5069	4	US-09-902-540-591	Sequence 591, App
C 36	8	3.8	5207	4	US-09-858-664A-1	Sequence 1, Appli
C 37	8	3.8	5207	4	US-10-274-978-1	Sequence 1, Appli
C 38	8	3.8	5207	4	US-10-274-978-3	Sequence 3, Appli
C 39	8	3.8	5207	4	US-10-697-263-1	Sequence 1, Appli
C 40	8	3.8	5207	4	US-10-697-263-3	Sequence 3, Appli
C 41	8	3.8	9663	4	US-09-869-588-22	Sequence 22, Appl
C 42	8	3.8	9726	4	US-09-869-588-21	Sequence 21, Appl
43	8	3.8	9813	4	US-09-949-016-13582	Sequence 13582, A
44	8	3.8	12047	2	US-09-022-461-1	Sequence 1, Appli
45	8	3.8	12047	3	US-09-033-556-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-949-016-5410  
; Sequence 5410, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5410  
; LENGTH: 1621  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-5410

Alignment Scores:  
Pred. No.: 1.17e-207  
Score: 209.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 4  
Length: 1621  
Matches: 209  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

AAH26175 (1-209) x US-09-949-016-5410 (1-1621)

Qy 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20

```
Db 24 ATGAATTCAGTTACATGTGAATCTTCCAGCGAATCAAGACCTGTGGTAATGAGT 83
Qy 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40
Db 84 GATGACAAACCCCTTCTATGTACTGCGCTGTGATGTGGCCACGCTTTTACCAACGAGAT 143
Qy 41 HisLeuAlaValHisIshisIshisGluMetThrLeuLysPheGlyProAlaArgAsn 60
Db 144 CATTTGGCTGTGCATAAACATAAATGAGATGACACTGAAATTTGGTCCAGCAGTAAT 203
Qy 61 AspSerValIleValAlaAspGlnThrProThrArgPheLeuLysAsnCysGlu 80
Db 204 GACAGTGTCTATGTGCTGATCAGACCCCAACCAACCAAGATTTCTTGAANAACCTGTGAA 263
Qy 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100
Db 264 GAAGTGGGTTTCTTTAATGAGTTGGGAGTCCATTGAGAAATGAATTCAGAAAGCTTCA 323
Qy 101 GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120
Db 324 GAAGATGACATTAATAAATGCTCTAGATTTATCCCTCTTGGCAACACCTATATCAAGA 383
Qy 121 SerLysIleGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140
Db 384 AGCAAAATTTGAGAGCTTCTGTGTGAGAAACAATCACAGATAGTCTTTTACTCTAC 443
Qy 141 ProGluSerThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr 160
Db 444 CCAGAGTCTACTACCACTGATCAGAGGAAGTACCATTGGCACAACCTGCACAGCCACA 503
Qy 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180
Db 504 TCAGCTATTGTTGCTCCAGCATCATACAGGTTCCCAATGTGCTGCTTACAAAGTTCTGAC 563
Qy 181 SerSerValIleIleGlnGlnAlaValProSerProThrSerSerThrValIleThrGln 200
Db 564 TCAAGTGTAAATTTATTCAGCAGCGAGTACCTTCACCAACCTCAAGTACTGTATATCACCAG 623
Qy 201 AlaProSerSerAsnArgProIleVal 209
Db 624 GCACCATCTCTTAACAGGCCAATTGTG 650

RESULT 2
US-09-949-016-17152
; Sequence 17152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17152
; LENGTH: 65966
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17152

Alignment Scores:
Pred. No.: 4,44e-51 Length: 65966
Score: 60.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
```

```
Query Match: 28.71% Indels: 0
DB: 4 Gaps: 0

AAH26175 (1-209) x US-09-949-016-17152 (1-65966)
Qy 150 GluValProLeuAlaGlnThrAlaGlnProThrSerAlaIleValArgProAlaSerLeu 169
Db 23601 GAAGTACCATTTGGCACAACCTGCACGCCACATCAGCTATGTTCGTCGAGCATCATTA 23660
Qy 170 GlnValProAsnValLeuLeuThrSerSerAspSerSerValIleIleGlnGlnAlaVal 189
Db 23661 CAGGTTCCCAATGTGCTCTTACAAGTTCTGACTCAAGTGTAATTTATTCAGCAGCAGTA 23720
Qy 190 ProSerProThrSerSerThrValIleThrGlnAlaProSerSerAsnArgProIleVal 209
Db 23721 CCTTACCAACCTCAAGTACTGTAAATCAACCCAGGACCACCTCTCTAACAGGCCAATTGTG 23780

RESULT 3
US-09-513-999C-2946
; Sequence 2946, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 2946
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 72..227
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 87
; OTHER INFORMATION: y=c or t
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 6
; OTHER INFORMATION: Xaa=Pro or Ser
US-09-513-999C-2946

Alignment Scores:
Pred. No.: 2.37e-23 Length: 229
Score: 31.00 Matches: 31
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.83% Indels: 0
DB: 4 Gaps: 0

AAH26175 (1-209) x US-09-513-999C-2946 (1-229)
Qy 29 AlaProGlyCysGlyGlnArgPheThrAsnGluAspHisLeuAlaValHisIshisLys 48
Db 102 GCGCCGGCTGTGGACAGAGATTTACAAACAGGAGCACCTGGCGAGTTCATAAACACCAAG 161
Qy 49 HisGluMetThrLeuLysPheGlyProAlaArg 59
Db 162 CATGATGATGACATTTGAAATTTGGCCACGCCGA 194

RESULT 4
US-09-949-016-189073/c
; Sequence 189073, Application US/09949016
; Patent No. 6812339
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 8, 2005, 14:03:03 ; Search time 530 Seconds  
(without alignments)  
2334.389 Million cell updates/sec

Title: AAH26175

Perfect score: 209

Sequence: 1 MKFKLVNSAQYKDLNMS.....PSPTSSTVITQAPSSNRPIV 209

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8776198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/AAH26175/runat\_07072005\_101846\_19958/app\_query.fasta.1.391  
-DB=N Geneseq 16Dec04 -CFMT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 @CGN 1 1 644 @runat\_07072005\_101846\_19958 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq\_16Dec04.\*

1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	1647	10 ADK67038	Adk67038 Gene #128
2	209	100.0	1647	10 ADK66978	Adk66978 Gene #68
3	144	68.9	3507	13 ADK08447	Adk08447 Full leng
4	132	63.2	1852	10 ADBS3173	Adbs3173 Primary r
5	58	27.8	3217	6 ABI99202	Abi99202 Mouse isc

6	57	27.3	501	6	AA661674	AA661674 Lung smal
7	34	16.3	509	10	ADD18510	ADD18510 Human pro
8	31	14.8	229	3	AAC02948	AAC02948 Human sec
9	31	14.8	800	5	ABV28993	ABV28993 Human pro
10	31	14.8	800	5	ABV23152	ABV23152 Human pro
11	31	14.8	1239	12	ADP08924	ADP08924 Mouse C13
12	31	14.8	2755	5	ABV28463	ABV28463 Human pro
13	31	14.8	2755	5	ABV22640	ABV22640 Human pro
14	19	9.1	2319	10	ADQ18732	ADQ18732 DNA of hu
15	19	9.1	2345	12	ADQ24589	ADQ24589 Human sof
16	19	9.1	2637	6	ABL69107	ABL69107 Kidney ca
17	19	9.1	2637	6	ABK83799	ABK83799 Human cDN
18	19	9.1	2637	10	ADH29017	ADH29017 Human chr
19	17	8.1	51	10	ADC22243	ADC22243 Protein b
20	16	7.7	51	10	ADC22231	ADC22231 Protein b
21	14	6.7	273	5	ABV04977	ABV04977 Human pro
22	13	6.2	397	5	ABV14146	ABV14146 Human pro
23	13	6.2	432	5	ABV44079	ABV44079 Human pro
24	13	6.2	432	5	ABV35240	ABV35240 Human pro
25	9	4.3	31	4	AAI30980	AAI30980 Human sin
26	9	4.3	527	4	AAK78581	AAK78581 Human imm
27	9	4.3	527	4	AAK78582	AAK78582 Human imm
28	9	4.3	1080	4	ABL21331	ABL21331 Drosophil
29	9	4.3	2775	5	AAH67959	AAH67959 C glutami
30	9	4.3	2775	8	ACA02052	ACA02052 C. glutam
31	9	4.3	2809	11	ADL65738	ADL65738 C. glutam
32	9	4.3	3199	4	ABL21330	ABL21330 Drosophil
33	9	4.3	9397	4	AAK78065	AAK78065 Human imm
34	9	4.3	254366	8	ABZ23704	ABZ23704 Human pho
35	9	4.3	349980	5	AAH68533	AAH68533 C glutami
36	8	3.8	139	8	ABX90792	ABX90792 Murine ge
37	8	3.8	158	3	AAQ22766	AAQ22766 Human sec
38	8	3.8	290	6	ABL73488	ABL73488 Corn tass
39	8	3.8	296	6	ABL85336	ABL85336 Human ova
40	8	3.8	333	6	ABZ34052	ABZ34052 HIV-1 rev
41	8	3.8	333	6	ABZ34065	ABZ34065 HIV-1 rev
42	8	3.8	410	4	AAI86431	AAI86431 Human pol
43	8	3.8	419	9	ACH21909	ACH21909 Human adu
44	8	3.8	431	4	ABK42999	ABK42999 Genomic s
45	8	3.8	431	4	ABK42998	ABK42998 Genomic s

#### ALIGNMENTS

RESULT 1  
ADK67038  
ID ADK67038 standard; DNA; 1647 BP.  
XX  
AC ADK67038;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Gene #128 for inhibitory RNA to manipulate stem cell phenotype.  
XX ds; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;  
XX pluripotent stem cell.  
XX Homo sapiens.  
XX WC2003068961-A2.  
XX  
XX 21-AUG-2003.  
XX  
XX 12-FEB-2003; 2003WO-GB000579.  
XX  
XX 13-FEB-2002; 2002GB-00003359.  
XX  
XX 13-FEB-2002; 2002GB-00003387.  
XX  
XX (AXOR-) AXORDIA LTD.  
XX  
XX Andrews P, Walsh J, Gokhale P;  
XX  
XX WPI; 2003-697528/66.

XX PT New inhibitory RNA molecule having double stranded RNA molecules, useful  
PT for manipulating the phenotype of stem cells, preferably pluripotent  
PT stem cells.  
XX  
XX  
XX Disclosure; SEQ ID NO 128; 157pp; English.  
XX  
XX The invention relates to an inhibitory RNA (RNAi) molecule derived from a  
CC nucleic acid molecule comprising a defined nucleic acid sequences given  
CC in the specification or a sequence which hybridizes to the sequences and  
CC encodes a Notch signaling target gene or which is a degenerate as a  
CC result of the genetic code of the sequences. The methods and compositions  
CC of the present invention are useful for manipulating the phenotype of  
CC stem cells, preferably pluripotent stem cells. This sequence corresponds  
CC to one of the nucleic acid molecules of the invention.  
XX  
SQ Sequence 1647 BP; 504 A; 396 C; 341 G; 406 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 1.01e-199 Length: 1647  
Score: 209.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0  
  
AAH26175 (1-209) x ADK67038 (1-1647)  
QY 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20  
Db 27 ATGAAATTCACGTTCATGTGAATCTGCCAGGCAATACAGGACCTGTGGAATATGAGT 86  
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40  
Db 87 GATGACAAACCCCTTTCTATGTACTCGCTGTGATGCGCCAGCGTTTACCAACGAGGAT 146  
QY 41 HisLeuAlaValHisLysHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60  
Db 147 CATTGGCTGTCCATAAATCAATGAGATGACACTGAAATTTGGTCCACGACGTAAAT 206  
QY 61 AspSerValIleValAlaAspGlnThrProThrArgPheLeuLysAsnCysGlu 80  
Db 207 GACAGTGTCTATTGGCTGTCAGACCCCAACCAACCAAGATTCTTGAAAAACTGTGAA 266  
QY 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100  
Db 267 GAAGTGGGTTCCTTTAATGAGTTGGCGAGTCCATTGGAATGAAATTCAGAAAGCTTCA 326  
QY 101 GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120  
Db 327 GAAGATGACATTAATAAATGCTCTAGATTATCCCTCTTGGCAACACCTATCATAGA 386  
QY 121 SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140  
Db 387 AGCAAAATTTAGGAGGCTTCCTGTGTAGAAACCACTCACCAGGATGCTTTACCTCAC 446  
QY 141 ProGluSerThrThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr 160  
Db 447 CCAGAGTCTACTACCATGATGAGAAGGAAGTACCATTGGCCACAACTGCACAGGCCACA 506  
QY 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180  
Db 507 TCAGCTATTGTCTGCCAGCATCATACAGGTTCCCAATGTGCTGCTTACAGTTCTGAC 566  
QY 181 SerSerValIleIleGlnGlnAlaValProSerProThrSerSerThrValIleThrGln 200  
Db 567 TCAAGTGTAAATTATTCAGCAGGACGATACCTTCACCAACCTCAAGTACTGTATATCACCAG 626  
QY 201 AlaProSerSerAsnArgProIleVal 209  
Db 627 GCACCATCTCTTAACAGGCCAATTGTGTC 653  
RESULT 2

ADK66978  
ID ADK66978 standard; DNA; 1647 BP.  
XX  
AC ADK66978;  
XX  
XX DT 06-MAY-2004 (first entry)  
XX  
XX Gene #68 for inhibitory RNA to manipulate stem cell phenotype.  
XX  
XX ds; gene; inhibitory RNA; RNAi; Notch signaling target gene; phenotype;  
KW pluripotent stem cell.  
XX  
XX Homo sapiens.  
XX  
XX WO2003068961-A2.  
XX  
XX PD 21-AUG-2003.  
XX  
XX PF 12-FEB-2003; 2003WO-GB000579.  
XX  
XX PR 13-FEB-2002; 2002GB-00003359.  
XX  
XX PR 13-FEB-2002; 2002GB-00003387.  
XX  
XX (AXOR-) AXORDIA LTD.  
XX  
XX PI Andrews P, Walsh J, Gokhale P;  
XX  
XX WIPI; 2003-697528/66.  
XX  
XX New inhibitory RNA molecule having double stranded RNA molecules, useful  
PT for manipulating the phenotype of stem cells, preferably pluripotent  
PT stem cells.  
XX  
XX Disclosure; SEQ ID NO 68; 157pp; English.  
XX  
XX The invention relates to an inhibitory RNA (RNAi) molecule derived from a  
CC nucleic acid molecule comprising a defined nucleic acid sequences given  
CC in the specification or a sequence which hybridizes to the sequences and  
CC encodes a Notch signaling target gene or which is a degenerate as a  
CC result of the genetic code of the sequences. The methods and compositions  
CC of the present invention are useful for manipulating the phenotype of  
CC stem cells, preferably pluripotent stem cells. This sequence corresponds  
CC to one of the nucleic acid molecules of the invention.  
XX  
SQ Sequence 1647 BP; 504 A; 396 C; 341 G; 406 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 1.01e-199 Length: 1647  
Score: 209.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0  
  
AAH26175 (1-209) x ADK66978 (1-1647)  
QY 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTrpAsnMetSer 20  
Db 27 ATGAAATTCACGTTCATGTGAATCTGCCAGGCAATACAGGACCTGTGGAATATGAGT 86  
QY 21 AspAspLysProPheLeuCysThrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40  
Db 87 GATGACAAACCCCTTTCTATGTACTCGCTGTGATGCGCCAGCGTTTACCAACGAGGAT 146  
QY 41 HisLeuAlaValHisLysHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60  
Db 147 CATTGGCTGTCCATAAATCAATGAGATGACACTGAAATTTGGTCCACGACGTAAAT 206  
QY 61 AspSerValIleValAlaAspGlnThrProThrArgPheLeuLysAsnCysGlu 80  
Db 207 GACAGTGTCTATTGGCTGTCAGACCCCAACCAACCAAGATTCTTGAAAAACTGTGAA 266  
QY 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: July 8, 2005, 14:05:28 ; Search time 4144 seconds  
(without alignments)  
2443.808 Million cell updates/sec

Title: AAH26175  
Perfect score: 209  
Sequence: 1 MKFKLVNSARQYKDLNMS.....PSSTSTVTIQAPSSNRPIV 209

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 9408497

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPto.spool/AAH26175/runat\_07072005\_101847\_19991/app\_query.fasta.1.391  
-DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=AAH26175 @CGN 1.1 4200 @runat\_07072005\_101847\_19991 -NCFU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	209	100.0	1370	9 BC026175	BC026175 Homo sapi
2	209	100.0	1647	6 AX821940	AX821940 Sequence
3	209	100.0	1647	6 AX822000	AX822000 Sequence
4	209	100.0	1647	9 HSCREBP1	X15875 Human mRNA

5	195	93.3	3957	10	BC042210	BC042210 Mus muscu
6	191	91.4	1464	10	AF483482	AF483482 Mus muscu
7	191	91.4	1464	10	AF483483	AF483483 Mus muscu
8	183	87.6	1596	10	S76657	S76657 cyclic AMP
9	144	68.9	3507	6	CQ851484	CQ851484 Sequence
10	144	68.9	3507	9	AK128731	AK128731 Homo sapi
11	143	68.4	1848	10	S76655	S76655 cyclic AMP
12	143	68.4	1848	10	HS0808617	HS0808617 Homo sapi
13	143	68.4	3983	9	BC079883	BC079883 Mus muscu
14	135	64.6	3909	10	BC082596	BC082596 Mus muscu
15	132	63.2	1209	10	RATRATF2	M65148 Rat RATF2 m
16	132	63.2	1852	10	RNU38938	U38938 Rattus norv
17	123	58.9	1302	10	S76659	S76659 cyclic AMP
18	106	50.7	1812	5	GGA17724	Y17724 Gallus gall
19	96	45.9	4851	12	AF050498	AF050498 Fusion tr
20	62	29.7	1520	5	XLU16158	U16158 Xenopus lae
21	60	28.7	176033	2	AL161794	AL161794 Homo sapi
22	60	28.7	178342	9	AC007435	AC007435 Homo sapi
23	60	28.7	207639	10	AL844581	AL844581 Mouse DNA
24	58	27.8	3217	6	AX305250	AX305250 Sequence
25	58	27.8	3217	10	MUSTCRATF	M77167 Mouse T-cel
26	58	27.8	166766	2	AC119682	AC119682 Rattus no
27	58	27.8	173793	2	AC129355	AC129355 Rattus no
28	57	27.3	501	6	AX321198	AX321198 Sequence
29	55	26.3	957	10	MUSMXP	M31629 Mouse mxBP
30	34	16.3	508	9	AY029364	AY029364 Homo sapi
31	33	15.8	11797	9	AC074291	AC074291 Homo sapi
32	31	14.8	229	6	AX887083	AX887083 Sequence
33	31	14.8	229	6	BD026693	BD026693 Sequence
34	31	14.8	800	6	CQ491274	CQ491274 Sequence
35	31	14.8	800	6	CQ491274	CQ491274 Sequence
36	31	14.8	1227	6	CQ720976	CQ720976 Sequence
37	31	14.8	1485	9	HSATFA1	X57197 H.sapiens m
38	31	14.8	2755	6	CQ490766	CQ490766 Sequence
39	31	14.8	2755	6	CQ496609	CQ496609 Sequence
40	31	14.8	2758	9	HSATFA	X52943 Human mRNA
41	31	14.8	2867	10	BC026483	BC026483 Mus muscu
42	29	13.9	787	9	BC042363	BC042363 Homo sapi
43	26	12.4	172558	9	AC023509	AC023509 Homo sapi
44	26	12.4	223132	10	AC123870	AC123870 Mus muscu
45	26	12.4	243444	2	AC133748	AC133748 Rattus no

#### ALIGNMENTS

RESULT 1  
BC026175  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BC026175 1370 bp mRNA linear PRI 16-SEP-2003  
Homo sapiens activating transcription factor 2, mRNA (cDNA clone IMAGE:4308091), complete cds.  
BC026175.1 GI:20072896  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1370)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Spatcheon,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., McSwan,P.J., Abramson,R.D., Mullah,J.A., Gunaratne,P.H., Richards,S., McKernan,K.J., Malek,J.A., Sodergren,E.J., Lu,X., Gibbs,R.A., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahy,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smallos, D.E., Schnerker, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257  
12477932  
2 (bases 1 to 1370)  
Straussberg, R.  
Direct Submission  
Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC/DCCTD/DTDP  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amad@systemsbiology.org](mailto:amad@systemsbiology.org)  
Anup Madan, Jessica Fahay, Erin Halton, Mark Kerteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 38 Row: c Column: 20  
This clone has the following problem: The cds is short compared to the longest cds in the locus.

## FEATURES

source  
1..1370  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4308091"  
/tissue\_type="Skin, melanotic melanoma, high MDR."  
/clone\_lib="NIH MGC\_62"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
1..1370  
/gene="ATF2"  
/note="synonyms: CRE-BP1, TREB7, HB16"  
/db\_xref="LocusID:1386"  
/db\_xref="MIM:123811"  
267..896  
/gene="ATF2"  
/codon\_start=1  
/product="ATF2 protein"  
/protein\_id="AAH26175.1"  
/db\_xref="GI:20072897"  
/db\_xref="LocusID:1386"  
/db\_xref="MIM:123811"  
/translation="MKFKLHVNSARQYKDLWNMSDDKFLCTAPCGORFTNEDHLAV HKHKEMILKGFARNSVIVADQPTPTFLKNCVEGLFNELASPFENBFKASD DIKMPDLSPLATIIRSKIEPVSQWETQDPSLPSTSDSKSVPLAQTAQPT SAIVRPASLIQVENLLTSSDSVIIQQAVPSPSTSVITQAPSSNRPIV"

## ORIGIN

Alignment Scores:  
Pred. No.: 1..968-209 Length: 1370  
Score: 209.00 Matches: 209  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

AAH26175 (1-209) x BC026175 (1-1370)

Qy 1 MetLysPheLysLeuHisValAsnSerAlaArgGlnTyrLysAspLeuTyrAsnMetSer 20  
|||||

Db 267 ATGAATTCAGTTTACATGTGTAATTCCTCCAGGCAATACAGGACCTCTGGAATATGAGT 326  
Qy 21 AspAspLysProPheLeuCythrAlaProGlyCysGlyGlnArgPheThrAsnGluAsp 40  
|||||  
Db 327 GATGACAAACCTTTCTATGTACTCGCGCTGGATGTGGCCAGCGTTTTTACCACAGGAT 386  
Qy 41 HisLeuAlaValHisLysHisGluMetThrLeuLysPheGlyProAlaArgAsn 60  
|||||  
Db 387 CATTTGGCTGTCCATAAACAATAGATGACACTGAAATTTGGTCCAGCAGTAAT 446  
Qy 61 AspSerValIleValAlaAspGlnThrProThrArgPheLeuLysAsnCysGlu 80  
|||||  
Db 447 GACAGTGTCTATGTGGCTGATCAGACCCCAACACCAACAGATTCTTGAAAACTGTGAA 506  
Qy 81 GluValGlyLeuPheAsnGluLeuAlaSerProPheGluAsnGluPheLysLysAlaSer 100  
|||||  
Db 507 GAAGTGGGTTTGTATTAATGAGTTGGCGAGTCCATTGAGAATGAATTCAGAAAGCTTCA 566  
Qy 101 GluAspAspIleLysLysMetProLeuAspLeuSerProLeuAlaThrProIleIleArg 120  
|||||  
Db 567 GAAGATGACATTAATAAATAATGCTCTAGATTATCCCTCTTGCACACCTATCATAGA 626  
Qy 121 SerLysIleGluGluProSerValValGluThrThrHisGlnAspSerProLeuProHis 140  
|||||  
Db 627 AGCAAAATTGAGGAGCCTTCTGTTGTAGAAACAACCTACACAGGATAGTCTTTACCTCAC 686  
Qy 141 ProGluSerThrThrSerAspGluLysGluValProLeuAlaGlnThrAlaGlnProThr 160  
Db 687 CCAGAGTCTACTACAGTGTGAGAGGAGTACCATTTGGCACAACCTGCACAGCCACA 746  
Qy 161 SerAlaIleValArgProAlaSerLeuGlnValProAsnValLeuLeuThrSerSerAsp 180  
|||||  
Db 747 TCAGCTATTGTTCGTCAGCATCATTCAGGTTCCTCAATGTCCTGCTACAGTTCTGCAC 806  
Qy 181 SerSerValIleIleGlnGlnAlaValProSerProThrSerSerThrValIleThrGln 200  
Db 807 TCAAGTGTAAATTATTCAGCAGCAGCAGTACCTTCACCAACCTCAAGTACTGTATCACCAG 866  
Qy 201 AlaProSerSerAsnArgProIleVal 209  
Db 867 GCACCATCTCTAACAGGCCAAATTGTG 893  
RESULT 2  
AX821940 1647 bp DNA linear PAT 10-DEC-2003  
LOCUS Sequence 68 from Patent WO03068961.  
DEFINITION AX821940  
ACCESSION AX821940  
VERSION AX821940.1 GI:39725161  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
AUTHORS  
TITLE  
JOURNAL  
METHOD to modify differentiation of pluripotential stem cells  
Patent: WO 03068961-A 68 21-AUG-2003;  
Axordia Limited (GB)  
FEATURES  
Location/Qualifiers  
1..1647  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2..31e-209 Length: 1647  
Score: 209.00 Matches: 209  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: July 8, 2005, 13:34:22 ; Search time 559 Seconds  
(without alignments)  
144.442 Million cell updates/sec

Title: AAH26175  
Perfect score: 209  
Sequence: 1 MKFKLVNSARQYKDLNMS.....PSPTSSTVITQAPSSNRPIV 209

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 1726216 seqs, 386330316 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1726216

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US10F\_PUBCOMB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	13.9	29	9	US-09-970-515-20
2	29	13.9	29	14	US-10-165-250A-20
3	29	13.9	29	15	US-10-342-683-20
4	29	13.9	29	15	US-10-457-614A-20
5	29	13.9	29	17	US-10-924-028-20
6	27	12.9	27	13	US-10-115-178-6
7	17	8.1	17	14	US-10-211-088-91
8	16	7.7	17	14	US-10-211-088-79
9	8	3.8	38	15	US-10-424-599-195921
10	8	3.8	125	16	US-10-425-115-232009
11	8	3.8	153	17	US-10-732-923-20872

12	8	3.8	523	16	US-10-437-963-129920	Sequence 129920,
13	8	3.8	951	16	US-10-437-963-167485	Sequence 167485,
14	8	3.8	1201	16	US-10-437-963-129457	Sequence 129457,
15	7	3.3	36	15	US-10-424-599-160301	Sequence 160301,
16	7	3.3	62	16	US-10-425-115-276081	Sequence 276081,
17	7	3.3	72	9	US-09-867-550-292	Sequence 292, App
18	7	3.3	74	15	US-10-424-599-256893	Sequence 256893,
19	7	3.3	78	15	US-10-424-599-212020	Sequence 212020,
20	7	3.3	79	15	US-10-424-599-145651	Sequence 145651,
21	7	3.3	100	16	US-10-425-115-276801	Sequence 276801,
22	7	3.3	101	16	US-10-425-115-285382	Sequence 285382,
23	7	3.3	101	16	US-10-425-115-351511	Sequence 351511,
24	7	3.3	115	16	US-10-425-115-398817	Sequence 398817,
25	7	3.3	120	9	US-09-738-626-6896	Sequence 6896, Ap
26	7	3.3	124	14	US-10-080-254-97	Sequence 97, Appl
27	7	3.3	124	15	US-10-242-355-641	Sequence 641, App
28	7	3.3	132	15	US-10-424-599-238800	Sequence 238800,
29	7	3.3	136	16	US-10-425-115-348784	Sequence 348784,
30	7	3.3	139	16	US-10-425-115-309456	Sequence 309456,
31	7	3.3	144	9	US-09-789-561-102	Sequence 102, App
32	7	3.3	144	11	US-09-833-245-2138	Sequence 2138, Ap
33	7	3.3	144	15	US-10-424-599-271504	Sequence 271504,
34	7	3.3	144	17	US-10-883-936-102	Sequence 102, App
35	7	3.3	147	16	US-10-425-115-320707	Sequence 320707,
36	7	3.3	147	16	US-10-425-115-358340	Sequence 358340,
37	7	3.3	151	15	US-10-424-599-257328	Sequence 257328,
38	7	3.3	153	15	US-10-425-114-66599	Sequence 66599, A
39	7	3.3	155	15	US-10-425-114-66627	Sequence 66627, A
40	7	3.3	159	15	US-10-425-114-63911	Sequence 63911, A
41	7	3.3	162	16	US-10-767-701-38653	Sequence 38653, A
42	7	3.3	164	16	US-10-437-963-198026	Sequence 198026,
43	7	3.3	174	9	US-09-839-185-12	Sequence 12, Appl
44	7	3.3	179	16	US-10-437-963-140436	Sequence 140436,
45	7	3.3	184	15	US-10-282-122A-52382	Sequence 52382, A

ALIGNMENTS

RESULT 1  
US-09-970-515-20  
; Sequence 20, Application US/09970515  
; Patent No. US20020127676A1  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; TITLE OF INVENTION: TRANSDUCTION PATHWAY  
; FILE REFERENCE: 20349-501 DIV  
; CURRENT APPLICATION NUMBER: US/09/970,515  
; PRIOR FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-09-970-515-20

Query Match 13.9%; Score 29; DB 9; Length 29;  
Best Local Similarity 100.0%; Pred. No. 2.5e+20;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 37 TNEHDHAVKHKHEWTLKFGPARNDVIV 65  
Db 1 TNEHDHAVKHKHEWTLKFGPARNDVIV 29

```
RESULT 2
US-10-165-250A-20
; Sequence 20, Application US/10165250A
; Publication No. US20030108539A1
; GENERAL INFORMATION:
; APPLICANT: Bonny, Christophe
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL
; FILE REFERENCE: 20349-501 CIP
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US/10/165,250A
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: 09/503,954
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: USSN 60/158,774
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically synthesized
US-10-165-250A-20

Query Match      13.9%; Score 29; DB 14; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.5e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEHDLAVHKHKHMTLKFGPARNDVIV 65
|||||
DB 1 TNEHDLAVHKHKHMTLKFGPARNDVIV 29

RESULT 3
US-10-342-683-20
; Sequence 20, Application US/10342683
; Publication No. US20030220480A1
; GENERAL INFORMATION:
; APPLICANT: Bonny, Christophe
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL
; FILE REFERENCE: 20349-501
; CURRENT APPLICATION NUMBER: US/10/342,683
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: US/09/503,954A
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: USSN 60/158,774
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically synthesized
US-10-342-683-20

Query Match      13.9%; Score 29; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.5e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEHDLAVHKHKHMTLKFGPARNDVIV 65
|||||
DB 1 TNEHDLAVHKHKHMTLKFGPARNDVIV 29

RESULT 4
US-10-457-614A-20
; Sequence 20, Application US/10457614A
; Publication No. US20040082509A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Bonny, Christophe
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL
; FILE REFERENCE: 20349-501CIP2
; CURRENT APPLICATION NUMBER: US/10/457,614A
; CURRENT FILING DATE: 2003-06-09
; PRIOR APPLICATION NUMBER: 60/158,774
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 09/503,954
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 10/165,250
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: 60/347,062
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically synthesized
US-10-457-614A-20

Query Match      13.9%; Score 29; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.5e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEHDLAVHKHKHMTLKFGPARNDVIV 65
|||||
DB 1 TNEHDLAVHKHKHMTLKFGPARNDVIV 29

RESULT 5
US-10-924-028-20
; Sequence 20, Application US/10924028
; Publication No. US20050043241A1
; GENERAL INFORMATION:
; APPLICANT: Bonny, Christophe
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL
; FILE REFERENCE: 20349-501 DIV
; CURRENT APPLICATION NUMBER: US/10/924,028
; CURRENT FILING DATE: 2004-08-23
; PRIOR APPLICATION NUMBER: US/09/970,515
; PRIOR FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 09/503,954
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: USSN 60/158,774
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically synthesized
US-10-924-028-20

Query Match      13.9%; Score 29; DB 17; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.5e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEHDLAVHKHKHMTLKFGPARNDVIV 65
|||||
DB 1 TNEHDLAVHKHKHMTLKFGPARNDVIV 29

RESULT 6
US-10-115-178-6
; Sequence 6, Application US/10115178
; Publication No. US20020119135A1
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 13:22:35 ; Search time 30 Seconds  
(without alignments)  
520.055 Million cell updates/sec

Title: AAH26175  
Perfect score: 209  
Sequence: 1 MKFKLVNSARQYKDLWNMS.....PSSTSTVITQAPSSNRPIV 209

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	501	4	US-09-949-016-11281
2	31	14.8	52	4	US-09-513-999C-7023
3	29	13.9	29	4	US-09-503-954A-20
4	29	13.9	29	4	US-09-970-515A-20
5	27	12.9	27	3	US-08-819-177-6
6	8	3.8	1053	4	US-09-328-352-5058
7	7	3.3	22	3	US-08-256-747C-1
8	7	3.3	22	3	US-08-834-130A-1
9	7	3.3	25	3	US-08-256-747C-57
10	7	3.3	25	3	US-08-834-130A-57
11	7	3.3	96	4	US-09-513-999C-7523
12	7	3.3	134	3	US-09-605-858-33
13	7	3.3	158	4	US-09-902-540-11525
14	7	3.3	193	4	US-09-248-796A-16279
15	7	3.3	196	4	US-09-949-016-11091
16	7	3.3	225	4	US-09-902-540-9773
17	7	3.3	229	5	PCT-US96-03916-13
18	7	3.3	229	5	PCT-US96-03916-62
19	7	3.3	258	4	US-09-485-529-20
20	7	3.3	396	3	US-09-134-001C-4546
21	7	3.3	405	4	US-09-248-796A-17965
22	7	3.3	418	4	US-09-710-279-380
23	7	3.3	430	3	US-09-134-001C-4302
24	7	3.3	447	4	US-09-248-796A-20950
25	7	3.3	481	1	US-07-927-071-1
26	7	3.3	481	6	5164481-1
27	7	3.3	481	6	5164481-1

28	7	3.3	690	4	US-09-538-092-1176	Sequence 1176, Ap
29	7	3.3	714	4	US-09-248-796A-19419	Sequence 19419, A
30	7	3.3	859	4	US-09-538-092-717	Sequence 717, App
31	7	3.3	992	1	US-08-127-499A-1	Sequence 1, Appli
32	7	3.3	992	1	US-08-482-847-1	Sequence 1, Appli
33	7	3.3	1063	1	US-08-093-453B-3	Sequence 3, Appli
34	7	3.3	1063	1	US-08-127-499A-8	Sequence 8, Appli
35	7	3.3	1063	1	US-08-482-847-8	Sequence 8, Appli
36	6	2.9	10	5	PCT-US95-03236-11	Sequence 11, Appli
37	6	2.9	17	3	US-09-308-022-3	Sequence 3, Appli
38	6	2.9	18	2	US-08-480-190-204	Sequence 204, App
39	6	2.9	18	2	US-08-488-379-204	Sequence 204, App
40	6	2.9	18	4	US-08-475-399A-204	Sequence 204, App
41	6	2.9	18	4	US-08-077-255A-204	Sequence 204, App
42	6	2.9	18	5	PCT-US93-07545-204	Sequence 204, App
43	6	2.9	20	2	US-08-934-915-8	Sequence 8, Appli
44	6	2.9	20	3	US-09-000-094-48	Sequence 48, Appli
45	6	2.9	20	4	US-10-011-743-48	Sequence 48, Appli

ALIGNMENTS

RESULT 1  
US-09-949-016-11281  
; Sequence 11281, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 11281  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-11281

Query Match	100.0%;	Score 209;	DB 4;	Length 501;
Best Local Similarity	100.0%;	Pred. No. 1.6e-193;		
Matches 209;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKFKLVNSARQYKDLWNMSDDKPFCTAPGCCQRTNEDHLAVHKHHEMTLKFGPARN	60	
Db	5	MKFKLVNSARQYKDLWNMSDDKPFCTAPGCCQRTNEDHLAVHKHHEMTLKFGPARN	64	
QY	61	DSVIVADQPTPTFRFLKNCNEVGLFNELASPPNEFKKASEDDIKMPLDLSPLATPIIR	120	
Db	65	DSVIVADQPTPTFRFLKNCNEVGLFNELASPPNEFKKASEDDIKMPLDLSPLATPIIR	124	
QY	121	SKIEEPSVETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQVNPVLLTSSD	180	
Db	125	SKIEEPSVETTHQDSPLPHPESTTSDEKEVPLAQTAQPTSAIVRPASLQVNPVLLTSSD	184	
QY	181	SSVITQAPVPSPTSSVTITQAPSSNRPIV	209	
Db	185	SSVITQAPVPSPTSSVTITQAPSSNRPIV	213	

RESULT 2  
US-09-513-999C-7023  
; Sequence 7023, Application US/09513999C  
; Patent No. 6783961

GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7023  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 6  
; OTHER INFORMATION: Xaa=Pro or Ser  
US-09-513-999C-7023

Query Match 14.8%; Score 31; DB 4; Length 52;  
Best Local Similarity 100.0%; Pred. No. 1.4e-22; Indels 0; Gaps 0;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AFGCGQFTNEDHLAVHKHKHMTLKFGPAR 59  
DB 11 AFGCGQFTNEDHLAVHKHKHMTLKFGPAR 41

RESULT 3  
US-09-503-954A-20  
; Sequence 20, Application US/09503954A  
; Patent No. 6610820  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501  
; CURRENT APPLICATION NUMBER: US/09/503,954A  
; CURRENT FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: USSN 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-09-503-954A-20

Query Match 13.9%; Score 29; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 7e-21; Indels 0; Gaps 0;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEDHLAVHKHKHMTLKFGPARNSVIV 65  
DB 1 TNEDHLAVHKHKHMTLKFGPARNSVIV 29

RESULT 4  
US-09-970-515A-20  
; Sequence 20, Application US/09970515A  
; Patent No. 6780970  
; GENERAL INFORMATION:  
; APPLICANT: Bonny, Christophe  
; TITLE OF INVENTION: CELL-PERMEABLE PEPTIDE INHIBITORS OF THE JNK SIGNAL  
; FILE REFERENCE: 20349-501 DIV

; CURRENT APPLICATION NUMBER: US/09/970,515A  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 09/503,954  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: 60/158,774  
; PRIOR FILING DATE: 1999-10-12  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically synthesized  
US-09-970-515A-20

Query Match 13.9%; Score 29; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 7e-21; Indels 0; Gaps 0;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 TNEDHLAVHKHKHMTLKFGPARNSVIV 65  
DB 1 TNEDHLAVHKHKHMTLKFGPARNSVIV 29

RESULT 5  
US-08-819-177-6  
; Sequence 6, Application US/08819177  
; Patent No. 6043083  
; GENERAL INFORMATION:  
; APPLICANT: Davis, Roger J.  
; APPLICANT: Dickens, Martin  
; TITLE OF INVENTION: INHIBITORS OF THE JNK SIGNAL  
; TITLE OF INVENTION: TRANSDUCTION PATHWAY AND METHODS OF USE  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/819,177  
; FILING DATE: 28 April 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. Peter  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/037001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION: ATF-2 JNK-binding domain  
US-08-819-177-6

Query Match 12.9%; Score 27; DB 3; Length 27;  
Best Local Similarity 100.0%; Pred. No. 5.6e-19; Indels 0; Gaps 0;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2005, 11:16:22 ; Search time 92 Seconds  
(without alignments)  
878.619 Million cell updates/sec

Title: AAH26175  
Perfect score: 209  
Sequence: 1 MKFKLVNSARQYKDLNWS.....PSPTSSTVITQAPSSNRPIV 209

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A Genesep\_16Dec04:.\*  
1: Genesep1980s.\*  
2: Genesep1990s.\*  
3: Genesep2000s.\*  
4: Genesep2001s.\*  
5: Genesep2002s.\*  
6: Genesep2003as.\*  
7: Genesep2003bs.\*  
8: Genesep2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	68.9	482	8	ADR10403 Human pro
2	58	27.8	411	5	ABBS7020 Mouse isc
3	31	14.8	52	3	AAG02942 Human sec
4	31	14.8	413	8	ADP80860 Mouse C13
5	29	13.9	29	4	AAB68325 c-Jun ami
6	29	13.9	29	7	ADP17925 Synthetic
7	29	13.9	29	8	ADO01349 c-Jun ami
8	19	9.1	501	7	ADC18733 Human cyc
9	17	8.1	17	7	ADC22242 Protein b
10	16	7.7	17	7	ADC22230 Protein b
11	9	4.3	359	4	ABB67482 Drosophil
12	8	3.8	95	3	AAG78911 Human pro
13	8	3.8	835	3	AAG31412 Arabidops
14	8	3.8	929	3	AAG31411 Arabidops
15	8	3.8	937	3	AAG31410 Arabidops
16	8	3.8	937	6	ABP59344 Arabidops
17	8	3.8	1053	6	ADA33771 Acinetoba
18	7	3.3	22	2	AAR39825 E1 peptid
19	7	3.3	25	2	AAR39879 Lipopeti
20	7	3.3	49	4	AAU31646 Novel hum
21	7	3.3	67	4	AAU91309 Human imm
22	7	3.3	72	5	ABP63776 Human ORF
23	7	3.3	83	4	AAO13691 Human pol
24	7	3.3	91	4	ABR10659 Human pan
25	7	3.3	91	4	AAU92701 Human dig

26	7	3.3	96	3	AAG03442 Human sec
27	7	3.3	99	4	AAU48473 Propionib
28	7	3.3	99	6	ABM44992 Propionib
29	7	3.3	103	2	AAR29888 HCV NS4-N
30	7	3.3	114	8	ADL10644 Human the
31	7	3.3	120	4	ADL10644 C glutami
32	7	3.3	124	4	AAM99904 Human exc
33	7	3.3	124	4	AAM43703 Human bla
34	7	3.3	124	8	ADF11662 Human bla
35	7	3.3	134	4	ABG1536 DegP prot
36	7	3.3	135	4	AAU45096 Propionib
37	7	3.3	135	6	ABM41615 Propionib
38	7	3.3	144	4	ABM7361 Human gen
39	7	3.3	144	5	ABG5389 Human alb
40	7	3.3	144	6	ABP58866 Tetrapept
41	7	3.3	144	8	ADL78656 Albumin f
42	7	3.3	158	8	ADN47355 Thermococ
43	7	3.3	165	4	AAO04300 Human pol
44	7	3.3	168	5	ABP7806 Human sec
45	7	3.3	168	6	ABP75874 Human sec

ALIGNMENTS

RESULT 1

ID	ADR10403	standard; protein; 482 AA.
XX	ADR10403;	
AC	ADR10403;	
XX	ADR10403;	
DT	04-NOV-2004	(first entry)
XX	04-NOV-2004	
DE	Human protein useful for treating neurological disease Seq 3909.	
XX	human; oligo-capping method; diagnostic marker; gene therapy;	
KW	osteoporosis; neurological disease; Alzheimer's disease;	
KW	Parkinson's disease; dementia; short memory; cancer;	
KW	sense or motor function; emotional reaction; fear response; panic;	
KW	osteopathic; neuroprotective; nootropic; antiparkinsonian; cytosstatic;	
KW	tranquilliser.	
XX	Homo sapiens.	
XX	EP1447413-A2.	
PN	EP1447413-A2.	
XX	EP1447413-A2.	
PD	18-AUG-2004.	
XX	18-AUG-2004.	
XX	12-FEB-2004; 2004EP-00003145.	
XX	12-FEB-2004; 2004EP-00003145.	
PR	14-FEB-2003; 2003JP-00102207.	
PR	03-MAY-2003; 2003JP-00131452.	
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.	
PA	Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;	
PI	Wakamatsu A, Ishii S, Nagai K, Irie R;	
PI	Wakamatsu A, Ishii S, Nagai K, Irie R;	
DR	WPI; 2004-583265/57.	
XX	N-PSDB; ADR08447.	
XX	New 1995 cDNA, useful for treating osteoporosis, neurological diseases,	
PT	Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.	
XX	Claim 1; SEQ ID NO 3909; 2686pp; English.	
CC	This invention relates to novel, isolated full length human cDNA	
CC	molecules and the encoded proteins thereof. Specifically, it refers to	
CC	cDNA clones obtained by an oligo-capping method, where none of these	
CC	clones are identical to any known human mRNAs. The present invention	
CC	describes an immunoassay to identify agonists and antagonists, as well as	
CC	antibodies, antisense molecules and siRNAs that can all be used to bind	
CC	to and modulate expression of the cDNA molecules. As such, these	

## BEST AVAILABLE COPY

CC molecules are useful for diagnostic markers or therapeutic targets for  
CC the various diseases or morbid states. In particular, they are useful in  
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
CC disease, Parkinson's disease, dementia, short memory and various cancers,  
CC as well as for maintaining equilibrium of sense or motor function, and  
CC for treating emotional reaction, fear response and panic. Accordingly,  
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
CC cytotstatic and tranquiliser activities. This polypeptide is a protein  
CC encoded by a full length human cDNA sequence of the invention. NOTE: This  
CC sequence is not given in the sequence listing of the specification but  
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-  
CC office.  
XX  
SQ

Sequence 482 AA;

Query Match 68.9%; Score 144; DB 8; Length 482;  
Best Local Similarity 100.0%; Pred. No. 2e-137;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ADQTPTRFLKNCBVEGLFNLASPFENEFKASEDDIKKMPDLSPPLATPIIRSKIEE 125  
DB 43 ADQTPTRFLKNCBVEGLFNLASPFENEFKASEDDIKKMPDLSPPLATPIIRSKIEE 102  
QY 126 PSVETHQDSPLPHPESTTSDEKEVPLAQTPSAIVRPASLOVPNVLLTSSDSSVII 185  
DB 103 PSVETHQDSPLPHPESTTSDEKEVPLAQTPSAIVRPASLOVPNVLLTSSDSSVII 162  
QY 186 QQAVPSPTSTVITQAPSSNRPIV 209  
DB 163 QQAVPSPTSTVITQAPSSNRPIV 186

## RESULT 2

ID ABB57020  
ID ABB57020 standard; protein; 411 AA.

AC ABB57020;

XX 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related protein sequence SEQ ID NO:2.

KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

OS Mus musculus.

XX WO200188188-A2.

XX 22-NOV-2001.

PF 18-MAY-2001; 2001WO-JP004192.

XX 18-MAY-2000; 2000JP-00145977.

XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

DR N-PSDB; ABI99202.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or by  
PT determining the expression profile of a gene group comprising these  
PT genes.

XX Claim 2; Page 51-53; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (I) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (I). The method

CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the  
CC protein sequences in ABB57020 to ABB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The expression  
CC levels or expression profiles produced by these genes are used as an  
CC indicator when screening for ischaemic condition-improving drugs or  
CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR  
CC primers for a mouse ischaemic condition related sequence, which are used  
CC in the exemplification of the present invention  
XX  
SQ Sequence 411 AA;

Query Match 27.8%; Score 58; DB 5; Length 411;  
Best Local Similarity 100.0%; Pred. No. 5.5e-50;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 PLAQTAPQTSIAIVRPASLOVPNVLLTSSDSSVIIQQAVPSPTSTVITQAPSSNRPIV 209  
DB 58 PLAQTAPQTSIAIVRPASLOVPNVLLTSSDSSVIIQQAVPSPTSTVITQAPSSNRPIV 115

## RESULT 3

AAG02942

ID AAG02942 standard; protein; 52 AA.

XX AAG02942;

AC 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7023.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.

OS Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122487P.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR N-PSDB; AAC02948.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 13; SEQ ID NO 7023; 71pp + Sequence Listing; English.

XX The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors  
XX  
SQ Sequence 52 AA;



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**This Page Blank (uspto)**